

# Estimation of a probability in inverse binomial sampling under normalized linear-linear and inverse-linear loss

Luis Mendo \*

December 4, 2013

## Abstract

Sequential estimation of the success probability  $p$  in inverse binomial sampling is considered in this paper. For any estimator  $\hat{p}$ , its quality is measured by the risk associated with normalized loss functions of linear-linear or inverse-linear form. These functions are possibly asymmetric, with arbitrary slope parameters  $a$  and  $b$  for  $\hat{p} < p$  and  $\hat{p} > p$  respectively. Interest in these functions is motivated by their significance and potential uses, which are briefly discussed. Estimators are given for which the risk has an asymptotic value as  $p \rightarrow 0$ , and which guarantee that, for any  $p \in (0, 1)$ , the risk is lower than its asymptotic value. This allows selecting the required number of successes,  $r$ , to meet a prescribed quality irrespective of the unknown  $p$ . In addition, the proposed estimators are shown to be approximately minimax when  $a/b$  does not deviate too much from 1, and asymptotically minimax as  $r \rightarrow \infty$  when  $a = b$ .

*Keywords:* Sequential estimation, Point estimator, Inverse binomial sampling, Asymmetric loss function.

## 1 Motivation and considered loss functions

The estimation of the success probability  $p$  of a sequence of Bernoulli trials is a recurring problem, arising in many branches of science and engineering. The quality of a point estimator of  $p$ , denoted as  $\hat{p}$ , can be measured in terms of its *risk*, or average loss associated with a certain *loss function*  $L$ . Since a given error is most meaningful when compared with the true value  $p$ , quality measures used in practice are most often normalized ones (Mendo, 2012). This corresponds to  $L$  being a function of  $\hat{p}/p$ , rather than of  $\hat{p}$ . Common loss functions include normalized squared error  $(\hat{p}/p - 1)^2$  and normalized absolute error  $|\hat{p}/p - 1|$ . Interval estimation can also be analyzed in terms of a certain loss function (Berger, 1985, p. 64) such that the resulting risk is the confidence level associated with an estimation interval.

Fixed-sample approaches to this problem suffer from the drawback that the required size depends on the unknown parameter  $p$ , and thus cannot be determined in advance. Therefore a *sequential procedure* is required, consisting of a *stopping rule*, which yields a random sample size, and an estimator based on the observed sample.

Sequential estimation in Bernoulli trials has been studied by many authors. Girshick et al (1946) introduce and analyze a general class of sequential procedures for

---

\*E.T.S. Ingenieros de Telecomunicación, Polytechnic University of Madrid, 28040 Madrid, Spain.  
E-mail: [lmendo@grc.ssr.upm.es](mailto:lmendo@grc.ssr.upm.es).

the unbiased estimation of  $p$ . Using a similar approach, DeGroot (1959) gives criteria for the selection of appropriate sampling plans for unbiased estimation of functions of  $p$ , with estimator performance measured by variance. His work shows that the fixed-size and inverse procedures are the only efficient sampling plans, i.e. the only ones for which the Cramer-Rao bound (or information inequality) holds with equal sign. Hubert and Pyke (2000) focus on the asymptotic behaviour as  $p \rightarrow 0$ . In this setting, they consider estimation of powers of  $p$ , with a loss function given as squared error loss multiplied by another power of  $p$ . Baran and Magiera (2010) carry out a similar asymptotic analysis for a different loss function. Using a Bayesian approach, Cabilio and Robbins (1975) and Cabilio (1977) consider symmetrized relative squared error loss (given as  $[(\hat{p}-p)/(p(1-p))]^2$ ) plus a fixed cost per observation, and find sequential procedures which minimize the Bayes risk in the estimation of  $p$ . Alvo (1977) studies sequential Bayes estimation from a more general point of view, where the observations are not necessarily Bernoulli variables, and considering squared error loss.

A particularly appealing stopping rule, first discussed by Haldane (1945), is *inverse binomial sampling* (also known as negative binomial sampling). Given  $r \in \mathbb{N}$ , this rule consists of taking as many observations as necessary to obtain exactly  $r$  successes. The random number of observations,  $N$ , is a sufficient statistic for  $p$  (Lehmann and Casella, 1998, p. 101). The interest in this stopping rule is motivated by the useful properties of the obtained estimators. Namely, it has been shown that for an estimator  $\hat{p} = g(N)$  such that  $\lim_{n \rightarrow \infty} ng(n)$  exists and is positive, and for a general class of loss functions defined by certain regularity conditions, the risk has an asymptotic value as  $p \rightarrow 0$  (Mendo, 2012). Moreover, estimators have been found whose risk for  $p$  arbitrary is guaranteed not to exceed its asymptotic value, for the specific cases of normalized mean squared error (Mikulski and Smith, 1976) (Sathe, 1977), normalized mean absolute error (Mendo, 2009) and confidence associated with a relative interval (Mendo and Hernando, 2006) (Mendo and Hernando, 2008) (Mendo and Hernando, 2010). This allows selecting an appropriate value of  $r$  that meets a prescribed risk irrespective of the unknown  $p$ .

In all cases mentioned in the preceding paragraph, the loss incurred by a negative error equals that of the corresponding positive error. In practice, however, situation-specific factors may render underestimation more or less costly than overestimation (Christoffersen and Diebold, 1997) (Akdeniz, 2004). Consider for example  $p = 0.01$  and two possible values of  $\hat{p}$ , namely 0.019 and 0.001. The absolute error (normalized or otherwise) is the same for both values of the estimator, as is the squared error. Nevertheless, with the first estimate  $\hat{p}$  is 1.9 times  $p$ , whereas with the second  $p$  is 10 times  $\hat{p}$ . In many applications it may be advisable to assign a higher loss to the second estimate. With absolute error, this could be accomplished by generalizing the loss function to one with a different slope on each side. Denoting  $x = \hat{p}/p$ , this generalized loss is given by

$$L(x) = \begin{cases} a(1-x) & \text{if } x \leq 1, \\ b(x-1) & \text{if } x > 1, \end{cases} \quad (1)$$

with parameters  $a, b \geq 0$ ,  $(a, b) \neq (0, 0)$ . This function, known as (normalized) *linear-linear* loss, frequently arises in applications; see for example Granger (1969) and Christoffersen and Diebold (1997). Another proposed function (not considered in this paper) which gives different weights to positive and negative errors is the *linear-exponential* loss, whose normalized version is  $L(x) = b[\exp(a(x-1)) - a(x-1) - 1]$ , with parameters  $a \neq 0$ ,  $b > 0$  (Akdeniz, 2004). The ratio  $a/b$ , in the linear-linear loss, or the parameter  $a$ , in the linear-exponential, control the relative importance given to underestimation and overestimation. Note that in both cases the loss due to

underestimation is bounded, unlike that of overestimation, which may be arbitrarily large.

In certain situations it may be meaningful to define loss as proportional to  $\hat{p}/p$  or  $p/\hat{p}$ , whichever is largest. Thus with the values in the previous example, the loss would be proportional to 1.9 and 10 respectively. In the following, the function  $s(\hat{p}, p) = \max\{\hat{p}/p, p/\hat{p}\}$  will be referred to as the *symmetric ratio* of  $\hat{p}$  and  $p$  (the name is motivated by the fact that  $s(\hat{p}, p) = s(p, \hat{p})$ ). The loss thus defined is inherently normalized, because it only depends on  $\hat{p}$  and  $p$  through  $x = \hat{p}/p$ . Subtracting 1 in order to have a minimum loss equal to 0, the loss function is expressed as  $L(x) = \max\{x, 1/x\} - 1$ . This function is unbounded for underestimation as well as for overestimation errors. In fact, its graph is symmetric about  $x = 1$  if  $\hat{p}$ , or  $x$ , is represented in logarithmic scale (this is obvious if  $L(x)$  is written as  $\exp|\log x| - 1$ ). The risk corresponding to this loss is the mean symmetric ratio minus 1, and represents a normalized measure of dissimilarity between  $\hat{p}$  and  $p$ , with smaller values corresponding to better estimators. A generalization is obtained, as before, by allowing different multiplicative parameters  $a, b \geq 0$ ,  $(a, b) \neq (0, 0)$  on each side of the function:

$$L(x) = \begin{cases} a(1/x - 1) & \text{if } x \leq 1, \\ b(x - 1) & \text{if } x > 1. \end{cases} \quad (2)$$

This will be referred to as *inverse-linear* loss.

The loss function (2), in addition to providing a natural measure of estimation quality, namely generalized mean symmetric ratio, can be representative of incurred cost in specific applications. In spite of this, it has not been used previously in the context of estimation problems, to the author's knowledge. As an example of application, consider the production of a certain device which is subject to manufacturing defects, such as image sensors for digital cameras. Several factors in the production process (such as the presence of dust particles) may result in a sensor with specific pixels systematically showing incorrect information. Since it would be too expensive to discard all sensors that have some defect, the commonly adopted solution is as follows. Each produced sensor is tested, and if the number of defective pixels is not too large it is accepted. The location of such pixels is permanently recorded in the camera, so that they can be corrected as a part of the processing applied by the camera to generate the image.

In high-quality camera models, however, it may be desirable to use sensors with an extremely low number of defects. A possible procedure is to classify each produced sensor as "premium" or "standard", depending on whether the number of pixel defects is extremely low or merely acceptable. Premium sensors are reserved for advanced cameras, which incorporate high-quality lenses, whereas standard sensors are mounted in consumer-level cameras with average-quality lenses. For ease of explanation, these two types of lenses will also be referred to as premium and standard, respectively. The production of each type of lens is a more deterministic process than that of sensors, and thus the number of produced lenses of each type is easily controlled.

It will be assumed that the manufacturer is primarily interested in its premium line of cameras. A number  $S$  of sensors is to be produced, and the amount of premium lenses that will be required needs to be planned in advance. To this end, an estimate  $\hat{p}$  is made of the proportion  $p$  of sensors that will turn out to be of the premium type (this can be done using inverse binomial sampling); and  $S\hat{p}$  premium lenses are made available. The actual proportion of premium sensors,  $p$ , may be lower than  $\hat{p}$ , in which case some of the premium lenses will be left unused; or it may be greater, and then some of the premium sensors will not be used. In either case, some resources are

wasted. If the cost associated with each unused part is  $a$  for a sensor and  $b$  for a lens, the risk computed from the loss function (2) is the average cost of wasted resources per assembled premium camera unit.

The rest of the paper analyzes inverse binomial sampling under the loss functions (1) and (2). The first has already been analyzed for the particular case  $a = b$  by Mendo (2009), and the generalization to  $a \neq b$  will be seen to be rather straightforward. The second function has not been dealt with before, to the author's knowledge, and its analysis turns out to be more difficult. Although the main focus of the paper is on the second, results for the first are also interesting by themselves. In each case, estimators are given in Section 2 such that the risk for  $p \in (0, 1)$  is guaranteed to be lower than its asymptotic value. Section 3 discusses these results and makes a comparison with the optimum performance that could be achieved by using other estimators. It is shown that the proposed estimators are approximately minimax if  $a/b$  is close to 1; and for  $a = b$  they are asymptotically minimax as  $r \rightarrow \infty$ . Section 4 contains the proofs to all results.

## 2 Main results

Consider a sequence of Bernoulli trials with probability of success  $p$ , and a random stopping time  $N$  given by inverse binomial sampling with  $r \in \mathbb{N}$ . Let  $x^{(i)}$  denote  $x(x-1)\cdots(x-i+1)$ , for  $x \in \mathbb{R}$ ,  $i \in \mathbb{N}$ ; and  $x^{(0)} = 1$ . The normalized lower incomplete gamma function is defined as

$$\gamma(t, u) = \frac{1}{\Gamma(t)} \int_0^u s^{t-1} \exp(-s) ds, \quad (3)$$

and satisfies the following well-known relationship (Abramowitz and Stegun, 1970, eq. (6.5.21)), which will be used throughout the paper:

$$\gamma(t-1, u) = \gamma(t, u) + \frac{u^{t-1} \exp(-u)}{\Gamma(t)}. \quad (4)$$

The random variable  $N$  has a negative binomial distribution, with probability function  $f_r(n) = P[N = n]$  given by  $f_r(n) = (n-1)^{(r-1)} p^r (1-p)^{n-r} / (r-1)!$ ,  $n \geq r$ . The corresponding distribution function will be denoted as  $F_r(n)$ . Similarly, the probability function of a binomial random variable with parameters  $n$  and  $p$  is denoted as  $b_{n,p}(i) = \binom{n}{i} p^i (1-p)^{n-i} / i!$ ,  $0 \leq i \leq n$ . For an arbitrary nonrandomized estimator  $\hat{p} = g(N)$  and a loss function  $L(\hat{p}/p)$ , the risk  $\eta(p)$  is

$$\eta(p) = E[L(\hat{p}/p)] = \sum_{n=r}^{\infty} f_r(n) L(g(n)/p). \quad (5)$$

For  $r \geq 2$  and  $a, b \geq 0$ , the loss functions (1) and (2) satisfy the sufficient conditions of Mendo (2012, theorem 1), and thus any estimator  $\hat{p} = g(N)$  with  $\lim_{n \rightarrow \infty} ng(n) = \Omega > 0$  has an asymptotic risk as  $p \rightarrow 0$ , which can be computed as

$$\lim_{p \rightarrow 0} \eta(p) = \frac{1}{(r-1)!} \int_0^{\infty} \nu^{r-1} \exp(-\nu) L(\Omega/\nu) d\nu. \quad (6)$$

In particular, this holds for any estimator that can be expressed as

$$\hat{p} = \frac{\Omega}{N + d} \quad (7)$$

with  $\Omega > 0$ ,  $d > -r$ .

Consider a generic estimator of the form (7). Denoting  $m = \lfloor \Omega/p - d \rfloor$ , the risk associated with the loss function (1) can be written as

$$\begin{aligned}\eta(p) &= a \sum_{n=m+1}^{\infty} \left(1 - \frac{\Omega}{(n+d)p}\right) f_r(n) + b \sum_{n=r}^m \left(\frac{\Omega}{(n+d)p} - 1\right) f_r(n) \\ &= -a \sum_{n=r}^{\infty} \left(\frac{\Omega}{(n+d)p} - 1\right) f_r(n) + (a+b) \sum_{n=r}^m \left(\frac{\Omega}{(n+d)p} - 1\right) f_r(n).\end{aligned}\quad (8)$$

Particularizing to  $\Omega = r - 1$  and  $d = -1$ , which yields the uniformly minimum variance unbiased (UMVU) estimator (Mikulski and Smith, 1976), and taking into account the identities (Mendo, 2009)

$$f_{r-1}(n-1) = \frac{(r-1)f_r(n)}{(n-1)p} \quad \text{for } r \geq 2, n \geq r, \quad (9)$$

$$F_{r-1}(n-1) = F_r(n) + (1-p)b_{n-1,p}(r-1) \quad \text{for } r \geq 2, n \geq r, \quad (10)$$

it is seen that, for  $r \geq 2$ , the first summand in (8) becomes 0, and

$$\begin{aligned}\eta(p) &= (a+b) \sum_{n=r}^m (f_{r-1}(n-1) - f_r(n)) = (a+b)(F_{r-1}(m-1) - F_r(m)) \\ &= (a+b)(1-p)b_{m-1,p}(r-1).\end{aligned}\quad (11)$$

The case  $a = b = 1$  is analyzed in Mendo (2009). Comparing (11) with Mendo (2009, eq. (12)), the expression of the risk for  $a, b$  arbitrary is seen to be a straightforward generalization of that for  $a = b = 1$ . As a consequence, the following result holds.

**Theorem 1.** *Consider the loss function given by (1) with  $a, b \geq 0$ ,  $(a, b) \neq (0, 0)$ . For  $r \geq 2$ , the risk  $\eta(p)$  associated with the estimator  $\hat{p} = (r-1)/(N-1)$  satisfies*

$$\eta(p) < \lim_{p \rightarrow 0} \eta(p) \quad \text{for any } p \in (0, 1), \quad (12)$$

with

$$\lim_{p \rightarrow 0} \eta(p) = \frac{(a+b)(r-1)^{r-2} \exp(-r+1)}{(r-2)!}. \quad (13)$$

In addition, as will be seen in Section 3, under certain conditions this estimator approaches the asymptotically optimum estimator discussed in Mendo (2012).

For the loss function (2), the risk associated with an estimator of the form (7) can be decomposed in a similar way as for (1). Namely,  $\eta(p) = \eta_1(p) + \eta_2(p)$  with

$$\eta_1(p) = a \sum_{n=m+1}^{\infty} \left(\frac{(n+d)p}{\Omega} - 1\right) f_r(n), \quad (14)$$

$$\eta_2(p) = b \sum_{n=r}^m \left(\frac{\Omega}{(n+d)p} - 1\right) f_r(n). \quad (15)$$

Assuming  $d \leq 0$  in (14) and taking into account that, as per (9),  $npf_r(n) = rf_{r+1}(n+1)$ , it follows that

$$\eta_1(p) \leq a \sum_{n=m+1}^{\infty} \left(\frac{np}{\Omega} - 1\right) f_r(n) = \frac{ar}{\Omega}(1 - F_{r+1}(m+1)) + a(F_r(m) - 1), \quad (16)$$

with strict inequality if  $d < 0$ . As for  $\eta_2(p)$ , assuming  $d \geq -1$ , it stems from (9) and (15) that

$$\eta_2(p) \leq b \sum_{n=r}^m \left( \frac{\Omega}{(n-1)p} - 1 \right) f_r(n) = \frac{b\Omega}{r-1} F_{r-1}(m-1) - bF_r(m), \quad (17)$$

with strict inequality if  $d > -1$  and  $m \geq r$ . As a result of (16) and (17), for any  $d \in [-1, 0]$  the risk satisfies

$$\eta(p) \leq \frac{b\Omega}{r-1} F_{r-1}(m-1) + (a-b)F_r(m) - \frac{ar}{\Omega} F_{r+1}(m+1) + a \left( \frac{r}{\Omega} - 1 \right). \quad (18)$$

The right-hand side of (18) is greatly simplified if  $\Omega$  is chosen as any value  $\tilde{\Omega} > 0$  such that

$$\frac{ar}{\tilde{\Omega}} - \frac{b\tilde{\Omega}}{r-1} = a - b, \quad (19)$$

for in that case, applying the identity (10),

$$\begin{aligned} \eta(p) &\leq \frac{b\tilde{\Omega}}{r-1} (F_{r-1}(m-1) - F_r(m)) + \frac{ar}{\tilde{\Omega}} (F_r(m) - F_{r+1}(m+1)) + a \left( \frac{r}{\tilde{\Omega}} - 1 \right) \\ &= \frac{b\tilde{\Omega}}{r-1} (1-p)b_{m-1,p}(r-1) + \frac{ar}{\tilde{\Omega}} (1-p)b_{m,p}(r) + a \left( \frac{r}{\tilde{\Omega}} - 1 \right). \end{aligned} \quad (20)$$

The advantage of this expression is that the terms  $(1-p)b_{m-1,p}(r-1)$  and  $(1-p)b_{m,p}(r)$  lend themselves to analysis more easily than the distribution functions in (18).

The condition (19) on  $\tilde{\Omega}$  has a single positive solution for  $a, b \geq 0$ ,  $(a, b) \neq (0, 0)$ , namely

$$\tilde{\Omega} = \begin{cases} (r-1) \left( 1 + \frac{a+b}{2b} \left( \sqrt{1 + \frac{4ab}{(r-1)(a+b)^2}} - 1 \right) \right) & \text{for } a, b > 0, \\ r-1 & \text{for } a = 0, b > 0, \\ r & \text{for } b = 0, a > 0. \end{cases} \quad (21)$$

It is easily seen that this reduces to  $\sqrt{r(r-1)}$  for  $a = b > 0$ . In addition, the following holds.

**Proposition 1.** *The value of  $\tilde{\Omega}$  given by (21) lies in the interval  $(r-1, r)$  for  $a, b > 0$ .*

As a consequence of Proposition 1, for any  $a, b \geq 0$ ,  $(a, b) \neq (0, 0)$ , the value  $\tilde{\Omega}$  defined by (21) satisfies  $\tilde{\Omega} \in [r-1, r]$ . Taking into account that  $\hat{p}_{\text{umvu}} = (r-1)/(N-1)$  is the UMVU estimator and that  $\hat{p}_{\text{ml}} = r/N$  is the maximum likelihood (ML) estimator (Best, 1974), the estimator  $\hat{p}$  given by (7) with  $\Omega \in [r-1, r]$  and  $d \in [-1, 0]$  is seen to be a ‘‘reasonable’’ one, in the sense that it is ‘‘close’’ to the UMVU and ML estimators. As will be seen in Section 3, in certain cases the proposed estimator is also close to the asymptotically optimum estimator in the sense of Mendo (2012).

The preceding arguments justify that the estimator given by (7) with  $\Omega \in [r-1, r]$  and  $d \in [-1, 0]$  is worth considering. In fact, for adequate choices of  $\Omega$  and  $d$ , it satisfies the important property that the risk is guaranteed not to exceed its asymptotic value, as established by the next theorem.

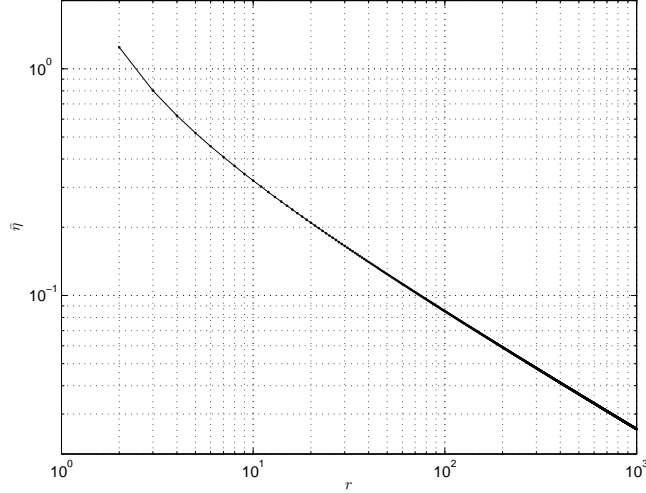


Figure 1: Risk guaranteed not to be exceeded for inverse-linear loss (2) with  $a = b = 1$

**Theorem 2.** Consider the loss function given by (2) with  $a, b \geq 0$ ,  $(a, b) \neq (0, 0)$ . For  $r \geq 2$ , the estimator  $\hat{p} = \tilde{\Omega}/N$  with  $\tilde{\Omega}$  given by (21) satisfies

$$\eta(p) < \lim_{p \rightarrow 0} \eta(p) \quad \text{for any } p \in (0, 1), \quad (22)$$

with

$$\lim_{p \rightarrow 0} \eta(p) = a \left( \frac{r}{\tilde{\Omega}} - 1 \right) + \left( \frac{ar}{\tilde{\Omega}} + b \right) \frac{\tilde{\Omega}^{r-1} \exp(-\tilde{\Omega})}{(r-1)!}. \quad (23)$$

## 3 Discussion and additional properties

### 3.1 Significance of the results

It has been shown in Section 2 that similar results to those already known for mean absolute error, mean squared error and confidence level also hold for generalized mean absolute error (Theorem 1) and generalized mean symmetric ratio (Theorem 2). Specifically, it has been proved that, for the proposed estimators,  $\sup_{p \in (0,1)} \eta(p) = \lim_{p \rightarrow 0} \eta(p)$ . In the following,  $\bar{\eta}$  will denote the value of  $\lim_{p \rightarrow 0} \eta(p)$ , or equivalently  $\sup_{p \in (0,1)} \eta(p)$ , for the estimators in Theorems 1 and 2.

The importance of these results lies in the fact that no knowledge is required about  $p$ . Thus, given any desired value  $\lambda$  for the risk, an adequate  $r$  can be selected such that the risk is guaranteed not to exceed  $\lambda$ , irrespective of  $p$ . Namely, it suffices to choose  $r$  as the minimum value for which  $\bar{\eta}$ , computed from (13) or from (23), is less than or equal to  $\lambda$ . As an illustration, Figure 1 depicts  $\bar{\eta}$  as a function of  $r$  for the loss given by (2) with  $a = b = 1$ . It is seen, for example, that  $r = 75$  suffices to guarantee a risk lower than 0.1, that is, a mean symmetric ratio lower than 1.1.

### 3.2 Comparison with minimax estimators

The presented results are valid for specific estimators, given by (7) with certain fixed values for  $\Omega$  and  $d$ . It is natural to ask to what extent the results could be improved

by considering other estimators, i.e. how much lower risks could be guaranteed not to be exceeded (or equivalently how much  $\sup_{p \in (0,1)} \eta(p)$  could be reduced). By definition, an estimator that is optimum according to this criterion (i.e. which minimizes  $\sup_{p \in (0,1)} \eta(p)$  over all possible estimators), if it exists, is a *minimax* estimator.

This question can be addressed on the basis of the analysis in Mendo (2012). For  $a, b > 0$ , both (1) and (2) satisfy the assumptions of Mendo (2012, theorem 3). This implies that there exists a value of  $\Omega$ , denoted as  $\Omega^*$ , such that any estimator  $\hat{p} = g(N)$  with  $\lim_{n \rightarrow \infty} ng(n) = \Omega^*$  minimizes  $\limsup_{p \rightarrow 0} \eta(p)$  over all estimators, including randomized ones. Thus any such estimator is asymptotically optimum, in the sense of achieving the minimum possible  $\limsup_{p \rightarrow 0} \eta(p)$ . This minimum, which will be denoted as  $\eta^*$ , restricts the values  $\lambda$  that the risk can be guaranteed not to exceed for  $p$  arbitrary. Namely, if an estimator guarantees that  $\eta(p) \leq \lambda$  for a given  $\lambda$ , then necessarily  $\lambda \geq \eta^*$ . As a consequence, the risk  $\bar{\eta}$  that is guaranteed not to be exceeded by the specific estimators considered in Section 2 is at most  $\bar{\eta}/\eta^*$  times larger than what could be achieved by a minimax estimator.

The value  $\eta^*$  is obtained as follows. Consider the loss function (1) first. For  $\Omega$  arbitrary, (6) gives

$$\lim_{p \rightarrow 0} \eta(p) = \frac{(a+b)\Omega\gamma(r-1, \Omega)}{r-1} - (a+b)\gamma(r, \Omega) + a \left(1 - \frac{\Omega}{r-1}\right). \quad (24)$$

Its derivative

$$\frac{d}{d\Omega} \lim_{p \rightarrow 0} \eta(p) = \frac{(a+b)\gamma(r-1, \Omega) - a}{r-1} \quad (25)$$

is seen to be monotone increasing. Therefore the minimizing value  $\Omega^*$  is unique, and is determined by the condition  $d(\lim_{p \rightarrow 0} \eta(p))/d\Omega = 0$ , that is,

$$\gamma(r-1, \Omega^*) = \frac{a}{a+b}. \quad (26)$$

Setting  $\Omega = \Omega^*$  in (24), substituting (26) and making use of (4),

$$\eta^* = \frac{(a+b)\Omega^{*r-1} \exp(-\Omega^*)}{(r-1)!}. \quad (27)$$

The value  $\Omega^*$  can be computed numerically from (26), and  $\eta^*$  is then obtained by means of (27).

Regarding the loss function (2), for  $\Omega$  arbitrary (6) gives

$$\lim_{p \rightarrow 0} \eta(p) = \frac{b\Omega\gamma(r-1, \Omega)}{r-1} + (a-b)\gamma(r, \Omega) - \frac{ar\gamma(r+1, \Omega)}{\Omega} + a \left(\frac{r}{\Omega} - 1\right). \quad (28)$$

Again, it is easily seen that

$$\frac{d}{d\Omega} \lim_{p \rightarrow 0} \eta(p) = \frac{b\gamma(r-1, \Omega)}{r-1} - \frac{ar(1 - \gamma(r+1, \Omega))}{\Omega^2} \quad (29)$$

is monotone increasing, and thus there is a single minimizing value  $\Omega^*$ , which satisfies

$$\frac{\Omega^{*2}}{r(r-1)} = \frac{a(1 - \gamma(r+1, \Omega^*))}{b\gamma(r-1, \Omega^*)}. \quad (30)$$

From (4), (28) and (30),

$$\eta^* = \left(a + \left(\frac{2\Omega^*}{r-1} - 1\right)b\right)\gamma(r-1, \Omega^*) + \frac{(b-a)\Omega^{*r-1} \exp(-\Omega^*)}{(r-1)!} - a. \quad (31)$$



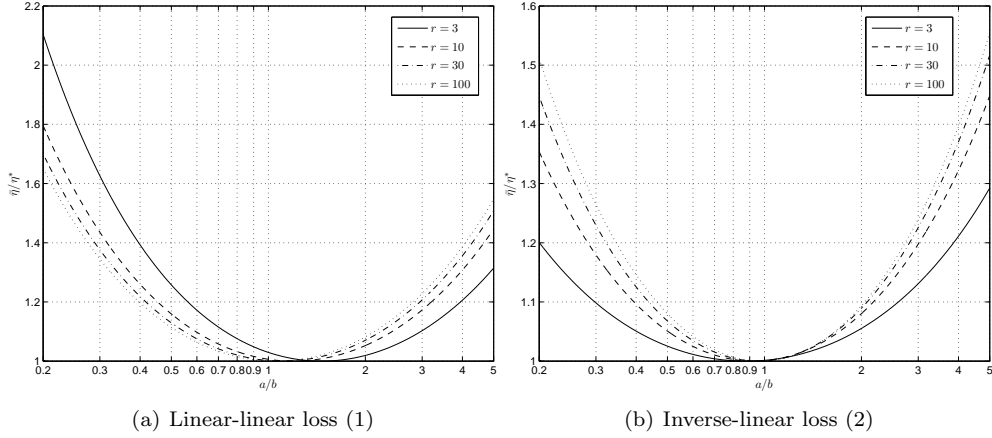


Figure 2: Degradation factor  $\bar{\eta}/\eta^*$  as a function of  $a/b$  and  $r$

The expressions (30) and (31) allow numerically computing  $\eta^*$ .

Figure 2 shows, for the loss functions and estimators considered in Theorems 1 and 2, the degradation factor  $\bar{\eta}/\eta^*$  as a function of  $a/b$ , with  $r$  as a parameter. As is seen, for  $a/b$  not too far from 1 the degradation factor is close to 1, that is, the considered estimators are nearly optimum. Furthermore, there is a value of  $a/b$  for which each estimator is precisely optimum, i.e. minimax, as established by the following.

**Proposition 2.** *For each of the loss functions (1) and (2), there exists a unique value of the ratio  $a/b$  for which the estimator considered in Theorem 1 or 2 respectively is minimax, that is, minimizes  $\sup_{p \in (0,1)} \eta(p)$  over all (possibly randomized) estimators. For the loss function (1) this value is given by*

$$\frac{a}{b} = \frac{\gamma(r-1, r-1)}{1 - \gamma(r-1, r-1)}, \quad (32)$$

and for (2) it is determined by the condition

$$\frac{\gamma(r-1, \tilde{\Omega})}{1 - \gamma(r+1, \tilde{\Omega})} = \frac{r(\tilde{\Omega} - r + 1)}{\tilde{\Omega}(r - \tilde{\Omega})} \quad (33)$$

with  $\tilde{\Omega}$  as in (21).

### 3.3 Minimacity for asymptotically large $r$ in the case $a = b$

The specific values of the ratio  $a/b$  determined by Proposition 2 can be shown to tend to 1 as  $r \rightarrow \infty$ . Related to this, the following establishes that for  $a = b$  the proposed estimators are asymptotically minimax as  $r \rightarrow \infty$ .

**Proposition 3.** *For the loss functions (1) and (2) with  $a = b$ , each of the estimators considered in Theorems 1 and 2, respectively, approaches a minimax estimator asymptotically as  $r \rightarrow \infty$ , in the sense that  $\lim_{r \rightarrow \infty} \bar{\eta}/\eta^* = 1$ .*

As a consequence of this result, for  $a = b$  and large  $r$  the considered estimators are approximately optimum in the minimax sense. This is illustrated in Figure 3, which

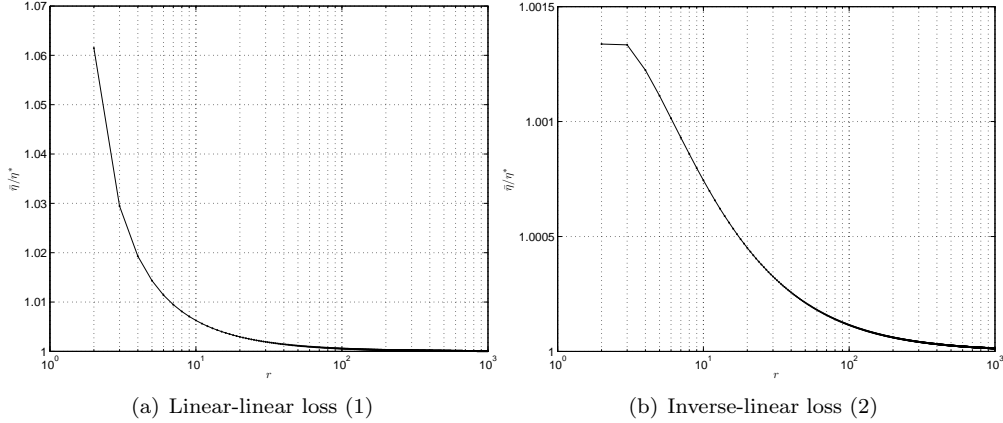


Figure 3: Degradation factor  $\bar{\eta}/\eta^*$  for  $a = b$  as a function of  $r$

shows the degradation factor  $\bar{\eta}/\eta^*$  as a function of  $r$ . In fact,  $\bar{\eta}/\eta^*$  is seen to be very low even for small  $r$ , and in particular for the range of values of  $r$  that are commonly used in practice. Thus, for example, the mean absolute error (loss function (1) with  $a = b = 1$ ) that is guaranteed not to be exceeded according to Theorem 1 is within 1% of the minimax mean absolute error for  $7 \leq r \leq 1000$ . Similarly, defining risk as mean symmetric ratio minus 1 (loss function (2) with  $a = b = 1$ ), the risk that is guaranteed not to be exceeded as per Theorem 2 is within 0.1% of the minimax risk for the same range of values of  $r$ .

## 4 Proofs

For  $\rho \in \mathbb{N}$ ,  $\rho \geq 2$ ;  $\mu \geq \rho - 1$ ; and  $t \in (0, 1)$ , let  $Y_\rho(\mu, t)$  be defined as

$$Y_\rho(\mu, t) = \frac{(\mu - 1)^{(\rho-1)} t^{\rho-1} (1-t)^{\mu-\rho+1}}{(\rho - 1)!}. \quad (34)$$

*Proof of Theorem 1.* The result immediately stems from (11) and the analysis in Mendo (2009).  $\square$   $\square$

**Lemma 1.** *The following inequality holds for  $r \geq 2$ ,  $d \in [-1, 0]$ ,  $j \geq 2$ .*

$$\sum_{i=1}^{r-1} (i + d + 1)^j > \frac{(r + d)^{j+1}}{j + 1}. \quad (35)$$

*Proof.* The sum in (35) can be expressed as the area covered by the  $r - 1$  rectangles of width 1 and height  $(i + d + 1)^j$ ,  $i = 1, \dots, r - 1$  in Figure 4, or equivalently as the shaded area comprised by  $r - 2$  unit-width trapezoids plus two half-width rectangles. Since the curve  $(x + d + 3/2)^j$  touches the upper vertices of the trapezoids and is convex, the following inequality can be written:

$$\begin{aligned} \sum_{i=1}^{r-1} (i + d + 1)^j &\geq \int_{1/2}^{r-3/2} \left(x + d + \frac{3}{2}\right)^j dx + \frac{(d+2)^j}{2} + \frac{(r+d)^j}{2} \\ &= \frac{(r+d)^{j+1}}{j+1} + \frac{(r+d)^j}{2} + (d+2)^j \left(\frac{1}{2} - \frac{d+2}{j+1}\right). \end{aligned} \quad (36)$$

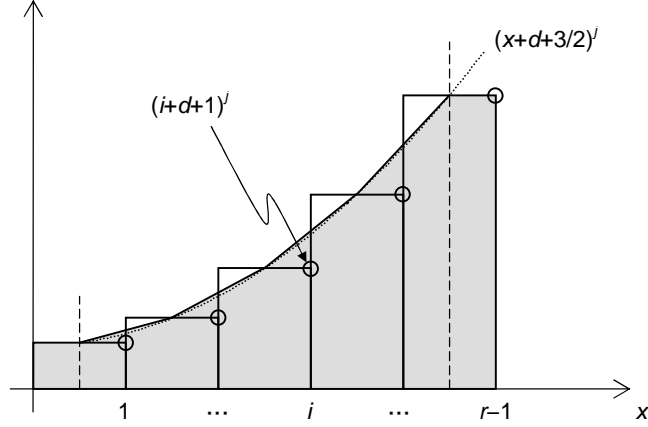


Figure 4: Illustration of (36)

For  $j \geq 3$  the term  $1/2 - (d+2)/(j+1)$  in (36) is nonnegative, which ensures that (35) holds. For  $j = 2$ , (36) reduces to

$$\sum_{i=1}^{r-1} (i+d+1)^j \geq \frac{(r+d)^3}{3} + \frac{-2d^3 - 6d^2 + 6(r-2)d + 3r^2 - 4}{6}. \quad (37)$$

The second summand in (37) has a derivative with respect to  $d$  equal to  $-d^2 - 2d + r - 2$ , which is nonnegative for  $d \in [-1, 0]$ ,  $r \geq 2$ . Thus this summand is lower bounded by its value at  $d = -1$ , i.e.  $(3r^2 - 6r + 4)/6$ , which is positive. Therefore (35) also holds for  $j = 2$ .  $\square$   $\square$

**Lemma 2.** *Given  $\rho$ ,  $\mu$ ,  $\Omega$  and  $\delta$  such that (i)  $\rho \in \mathbb{N}$ ,  $\rho \geq 2$ ; (ii)  $\Omega \in [\rho - 1, \rho]$ ; (iii)  $\delta \in [-1, 0]$ ; (iv)  $\mu > \rho - 1$ ; and (v)  $\mu > \Omega - \delta - 1$ , the following hold:*

(a)  $Y_\rho(\mu, \Omega/(\mu + \delta + 1))$  is a strictly increasing function of  $\mu$ , with

$$\lim_{\mu \rightarrow \infty} Y_\rho(\mu, \Omega/(\mu + \delta + 1)) = \Omega^{\rho-1} \exp(-\Omega)/(\rho - 1)!. \quad (38)$$

(b)  $Y_{\rho+1}(\mu + 1, \Omega/(\mu + \delta + 1))$  is a strictly increasing function of  $\mu$ , with

$$\lim_{\mu \rightarrow \infty} Y_{\rho+1}(\mu + 1, \Omega/(\mu + \delta + 1)) = \Omega^\rho \exp(-\Omega)/\rho!. \quad (39)$$

*Proof.* According to hypotheses (iv) and (v), it holds that  $\mu > \rho - 1$  and  $\Omega/(\mu + \delta + 1) < 1$ , and thus  $Y_\rho(\mu, \Omega/(\mu + \delta + 1))$  and  $Y_{\rho+1}(\mu + 1, \Omega/(\mu + \delta + 1))$  are well defined from (34).

The proof will be carried out separately for parts (a) and (b) of the Lemma.

(a) It is convenient to make the change of variable  $t = \Omega/(\mu + \delta + 1)$ , by which  $Y_\rho(\mu, \Omega/(\mu + \delta + 1))$  is expressed as  $Y_\rho(\Omega/t - \delta - 1, t)$ . It will be shown that

$$\lim_{t \rightarrow 0} Y_\rho(\Omega/t - \delta - 1, t) = \Omega^{\rho-1} \exp(-\Omega)/(\rho - 1)!, \quad (40)$$

which is equivalent to (38); and that  $Y_\rho(\Omega/t - \delta - 1, t)$  is a strictly decreasing function of  $t$ , which will imply that  $Y_\rho(\mu, \Omega/(\mu + \delta + 1))$  strictly increases with  $\mu$ . From (34),

$$\log Y_\rho(\mu, t) = \sum_{i=1}^{\rho-1} \log \frac{(\mu - i)t}{\Omega} + \sum_{i=1}^{\rho-1} \log \frac{\Omega}{i} + (\mu - \rho + 1) \log(1 - t), \quad (41)$$

and thus

$$\begin{aligned} \log Y_\rho \left( \frac{\Omega}{t} - \delta - 1, t \right) &= \sum_{i=1}^{\rho-1} \log \left( 1 - \frac{(i + \delta + 1)t}{\Omega} \right) \\ &+ \sum_{i=1}^{\rho-1} \log \frac{\Omega}{i} + \left( \frac{\Omega}{t} - \rho - \delta \right) \log(1 - t). \end{aligned} \quad (42)$$

Taking into account that  $(\rho + \delta)t/\Omega = (\rho + \delta)/(\mu + \delta + 1) < 1$  as a result of (iv), and that  $t < 1$ , the Taylor expansion  $\log(1 - t) = -\sum_{j=1}^{\infty} t^j/j$ ,  $|t| < 1$  can be used in (42) to yield

$$\log Y_\rho \left( \frac{\Omega}{t} - \delta - 1, t \right) = -\sum_{j=0}^{\infty} c_j t^j, \quad (43)$$

$$c_0 = \Omega + \sum_{i=1}^{\rho-1} \log \frac{i}{\Omega}, \quad (44)$$

$$c_j = \frac{\Omega}{j+1} - \frac{\rho + \delta}{j} + \frac{1}{j\Omega^j} \sum_{i=1}^{\rho-1} (i + \delta + 1)^j \quad \text{for } j \geq 1. \quad (45)$$

The equalities (43) and (44) imply (40), and thus (38).

To prove that  $Y_\rho(\Omega/t - \delta - 1, t)$  strictly decreases with  $t$ , it suffices to show that the coefficients  $c_j$  satisfy  $c_j \geq 0$ ,  $j \geq 1$ , with strict inequality for some  $j$ . For  $j \geq 2$ , (45) and Lemma 1 yield

$$\begin{aligned} j(j+1)c_j &> j\Omega - (j+1)(\rho + \delta) + (\rho + \delta) \left( \frac{\rho + \delta}{\Omega} \right)^j \\ &= -j(\rho + \delta - \Omega) + (\rho + \delta) \left( \left( \frac{\rho + \delta}{\Omega} \right)^j - 1 \right). \end{aligned} \quad (46)$$

Taking into account that  $\rho + \delta \geq 0$  by hypothesis (iii), and using the inequality

$$\left( \frac{\rho + \delta}{\Omega} \right)^j = \left( 1 + \frac{\rho + \delta - \Omega}{\Omega} \right)^j \geq 1 + \frac{j(\rho + \delta - \Omega)}{\Omega}, \quad (47)$$

it follows from (46) that

$$c_j > \frac{(\rho + \delta - \Omega)^2}{(j+1)\Omega} \geq 0. \quad (48)$$

For  $j = 1$ , (45) gives

$$c_1 = \frac{\Omega^2 - 2\rho\Omega + (\rho - 1)(\rho + 2)}{2\Omega} + \frac{\delta(\rho - \Omega - 1)}{\Omega}. \quad (49)$$

Consider the first summand in (49). The minimum of its numerator with respect to  $\Omega$  is attained at  $\Omega = \rho$  and equals  $\rho - 2$ . Thus, according to hypothesis (i), this summand is nonnegative. By (ii) and (iii), the second summand is also nonnegative; and therefore  $c_1 \geq 0$ . Consequently  $Y_\rho(\Omega/t - \delta - 1, t)$  strictly decreases with  $t$ , and thus  $Y_\rho(\mu, \Omega/(\mu + \delta + 1))$  strictly increases with  $\mu$ .

(b) Making the same change of variable as in part (a), and taking into account (43)–(45),

$$\log Y_{\rho+1} \left( \mu + 1, \frac{\Omega}{\mu + \delta + 1} \right) = \log Y_{\rho+1} \left( \frac{\Omega}{t} - \delta, t \right) = - \sum_{j=0}^{\infty} c'_j t^j, \quad (50)$$

$$c'_0 = \Omega + \sum_{i=1}^{\rho} \log \frac{i}{\Omega}, \quad (51)$$

$$c'_j = \frac{\Omega}{j+1} - \frac{\rho + \delta}{j} + \frac{1}{j\Omega^j} \sum_{i=0}^{\rho-1} (i + \delta + 1)^j \geq c_j \quad \text{for } j \geq 1. \quad (52)$$

From (52) it follows that  $c'_1 \geq 0$  and  $c'_j > 0$  for  $j \geq 2$ . Together with (50) and (51), this establishes part (b) of the Lemma.  $\square$   $\square$

*Proof of Theorem 2.* As the case  $a = 0, b > 0$  is already covered by Theorem 1, it will be assumed that  $a > 0$ . This implies, according to Proposition 1, that  $\tilde{\Omega} > r - 1$ .

The equality (23) is obtained substituting the loss function (2) into (6) with  $\Omega = \tilde{\Omega}$ , and making use of (4) and (19).

The inequality (20) can be expressed as

$$\eta(p) \leq \frac{b\tilde{\Omega}Y_r(m,p)}{r-1} + \frac{arY_{r+1}(m+1,p)}{\tilde{\Omega}} + a \left( \frac{r}{\tilde{\Omega}} - 1 \right), \quad (53)$$

$$m = \lfloor \tilde{\Omega}/p \rfloor. \quad (54)$$

From (54) it stems that  $m \geq r - 1$ . Each value of  $m$  has an associated interval  $I_m \subseteq (0, 1)$  such that (54) holds if and only if  $p \in I_m$ . Namely,  $I_m = (p_l, p_u]$  with  $p_l = \tilde{\Omega}/(m+1)$ ,  $p_u = \tilde{\Omega}/m$ , except if  $m = r - 1$ , in which case  $\tilde{\Omega} < r$  and thus  $p_l < 1$ ,  $p_u > 1$ ; or if  $m = r$  and  $\tilde{\Omega} = r$ , which gives  $p_l < 1$ ,  $p_u = 1$ ; in either case  $I_m = (p_l, 1)$ . According to (53), and taking into account (19), to establish (22) it suffices to show that, for  $p \in (0, 1)$  and  $m$  given by (54),

$$Y_r(m,p) < \tilde{\Omega}^{r-1} \exp(-\tilde{\Omega})/(r-1)!, \quad (55)$$

$$Y_{r+1}(m+1,p) < \tilde{\Omega}^r \exp(-\tilde{\Omega})/r!. \quad (56)$$

If  $m = r - 1$  the left-hand sides of (55) and (56) are zero, and the inequalities are clearly satisfied. Thus in the following it will be assumed that  $m \geq r$ .

As a step in the proof of (55) and (56), it will be shown that for  $p \in (0, 1)$  and  $m \geq r$  related by (54), or equivalently for  $m \geq r$  and  $p \in I_m$ , the following inequalities hold:

$$Y_r(m,p) \leq \begin{cases} Y_r(m, (r-1)/m) & \text{if } (\tilde{\Omega} - r + 1)m \leq r - 1, \\ Y_r(m, \tilde{\Omega}/(m+1)) & \text{if } (\tilde{\Omega} - r + 1)m > r - 1. \end{cases} \quad (57)$$

$$Y_{r+1}(m+1,p) \leq \begin{cases} Y_{r+1}(m+1, r/(m+1)) & \text{if } (r - \tilde{\Omega})m \leq \tilde{\Omega}, \\ Y_{r+1}(m+1, \tilde{\Omega}/m) & \text{if } (r - \tilde{\Omega})m > \tilde{\Omega}. \end{cases} \quad (58)$$

For  $m \geq r$ , it follows from (34) that  $Y_r(m,p)$  considered as a function of  $p \in (0, 1)$  is maximum at  $p_{\max} = (r-1)/m < 1$ , monotone increasing for  $p < p_{\max}$ , and monotone decreasing for  $p > p_{\max}$ . As  $\tilde{\Omega} > r - 1$ , it is seen that  $p_{\max} < p_u \leq 1$ , and that

$p_{\max} < p_l$  if and only if  $(\tilde{\Omega} - r + 1)m > r - 1$ . This implies that  $Y_r(m, p)$  is bounded as given by (57). Regarding (58), the maximum of  $Y_{r+1}(m + 1, p)$  with respect to  $p \in (0, 1)$  is attained at  $p'_{\max} = r/(m + 1) < 1$ . As  $m \geq r$ , it stems that  $p_l \leq p'_{\max} < 1$ , and that  $p_u < p'_{\max}$  if and only if  $(r - \tilde{\Omega})m > \tilde{\Omega}$ . This establishes (58).

The proof of (55) will be based on (57). Since  $\tilde{\Omega} > r - 1$ , the following definition can be made:  $\mu_1 = (r - 1)/(\tilde{\Omega} - r + 1)$ . The fact that  $\tilde{\Omega} \leq r$  implies that  $\mu_1 \geq r - 1$ . The upper condition in (57) is equivalent to  $m \leq \mu_1$ , whereas the lower corresponds to  $m > \mu_1$ . As  $m$  cannot be smaller than  $r$ , the condition  $m \leq \mu_1$  can only be met for some  $m$  if  $\mu_1 \geq r$ , i.e. if  $\tilde{\Omega} \leq r - 1/r$ . On the other hand, the condition  $m > \mu_1$  can always be satisfied by taking  $m$  sufficiently large. Thus, (55) will be established in two steps. First, it will be shown that  $Y_r(m, \tilde{\Omega}/(m + 1))$  monotonically increases with  $m > \mu_1$  and tends to  $\tilde{\Omega}^{r-1} \exp(-\tilde{\Omega})/(r - 1)!$  as  $m \rightarrow \infty$ . This will prove that (55) holds for all  $m > \mu_1$ . Second, it will be shown, for  $\mu_1 \geq r$ , that  $Y_r(m, (r - 1)/m)$  monotonically increases with  $m \geq r$  and is smaller than  $\tilde{\Omega}^{r-1} \exp(-\tilde{\Omega})/(r - 1)!$  for  $m = \mu_1$ . This will establish (55) for all  $m$  such that  $r \leq m \leq \mu_1$ .

Regarding the first case,  $m > \mu_1$ , consider Lemma 2(a) with values  $r, m, \tilde{\Omega}, 0$  respectively for  $\rho, \mu, \Omega, \delta$ . These values satisfy the hypotheses of the Lemma (it is obvious that (i)–(iii) hold; (iv) and (v) are satisfied as well because  $m > \mu_1 \geq r - 1 \geq \tilde{\Omega} - 1$ ). According to this,  $Y_r(m, \tilde{\Omega}/(m + 1))$  monotonically increases with  $m$  and tends to  $\tilde{\Omega}^{r-1} \exp(-\tilde{\Omega})/(r - 1)!$  as  $m \rightarrow \infty$ . Therefore (55) holds for  $m > \mu_1$ .

For the case  $r \leq m \leq \mu_1$ ,  $\mu_1 \geq r$ , using Lemma 2(a) (with values  $r, m, r - 1, -1$  respectively for  $\rho, \mu, \Omega, \delta$ ; (iv) and (v) hold because  $m \geq r$ ) it is seen that  $Y_r(m, (r - 1)/m)$  increases with  $m$ . The definition of  $\mu_1$  implies that  $(r - 1)/\mu_1 = \tilde{\Omega}/(\mu_1 + 1)$ , and thus

$$Y_r(\mu_1, (r - 1)/\mu_1) = Y_r(\mu_1, \tilde{\Omega}/(\mu_1 + 1)). \quad (59)$$

Applying Lemma 2(a) again (with values  $r, \mu_1, \tilde{\Omega}, 0$ ; note that (v) is satisfied because  $\mu_1 \geq r \geq \tilde{\Omega} > \tilde{\Omega} - 1$ ) to the right-hand side of this equality shows that (59) is smaller than  $\tilde{\Omega}^{r-1} \exp(-\tilde{\Omega})/(r - 1)!$ . Therefore (55) holds for  $r \leq m \leq \mu_1$ .

As for (56), it is seen that the lower condition in (58) is not met for any  $m$  if  $\tilde{\Omega} = r$ , whereas if  $\tilde{\Omega} < r$  there exist values of  $m$  which satisfy each of the conditions. These two cases will be treated separately.

In the case  $\tilde{\Omega} < r$ , the proof proceeds along the same lines as that of (55). Let  $\mu_2 = \tilde{\Omega}/(r - \tilde{\Omega})$ . The fact that  $\tilde{\Omega} > r - 1$  implies that  $\mu_2 > r - 1$ . In addition, since  $m \geq r$ , the upper condition in (58) can only be met if  $\mu_2 \geq r$ . Thus it suffices to show first that  $Y_{r+1}(m + 1, \tilde{\Omega}/m)$  monotonically increases with  $m > \mu_2$  and tends to  $\tilde{\Omega}^r \exp(-\tilde{\Omega})/r!$  as  $m \rightarrow \infty$ ; and second that, if  $\mu_2 \geq r$ ,  $Y_{r+1}(m + 1, r/(m + 1))$  monotonically increases with  $m \geq r$  and is smaller than  $\tilde{\Omega}^r \exp(-\tilde{\Omega})/r!$  for  $m = \mu_2$ . The first part directly stems from Lemma 2(b) (with values  $r, m, \tilde{\Omega}, -1$ ). As for the second, the increasing character of  $Y_{r+1}(m + 1, r/(m + 1))$  with  $m$  is also established by Lemma 2(b) (with values  $r, m, r, 0$ ). The definition of  $\mu_2$  implies that  $r/(\mu_2 + 1) = \tilde{\Omega}/\mu_2$ , from which

$$Y_{r+1}(\mu_2 + 1, r/(\mu_2 + 1)) = Y_{r+1}(\mu_2 + 1, \tilde{\Omega}/\mu_2), \quad (60)$$

and applying Lemma 2(b) (with values  $r, \mu_2, \tilde{\Omega}, -1$ ; (iv) and (v) hold because  $\mu_2 = \tilde{\Omega}/(r - \tilde{\Omega}) > \tilde{\Omega} > r - 1$ ) to the right-hand side of (60) establishes that it is smaller than  $\tilde{\Omega}^r \exp(-\tilde{\Omega})/r!$ .

In the case  $\tilde{\Omega} = r$ , the expression (58) reduces to its upper part, and (56) follows from Lemma 2(b) (with values  $r, m, r, 0$ ). This completes the proof.  $\square$   $\square$

*Proof of Proposition 2.* For  $L$  as in (1), equating  $d(\lim_{p \rightarrow 0} \eta(p))/d\Omega$  given by (25) to 0, solving for  $a/b$  and particularizing to  $\Omega = r - 1$  yields (32).

As for  $L$  given by (2), from (19) it is seen that

$$\frac{a}{b} = \frac{(\tilde{\Omega} - r + 1)\tilde{\Omega}}{(r - 1)(r - \tilde{\Omega})}. \quad (61)$$

Setting  $\Omega^* = \tilde{\Omega}$  in (30) and combining with (61) yields (33).  $\square$   $\square$

**Lemma 3.** *For any  $k \in \mathbb{N}$ , the factorial  $k!$  satisfies the following:*

$$k! > \sqrt{2\pi} k^{k+1/2} \exp(-k), \quad (62)$$

$$\lim_{k \rightarrow \infty} \frac{k! \exp(k)}{k^{k+1/2}} = \sqrt{2\pi}. \quad (63)$$

*Proof.* These expressions follow from Abramowitz and Stegun (1970, eq. (6.1.38)).  $\square$   
 $\square$

**Lemma 4.** *For any sequence of numbers  $\delta_k$  such that  $0 \leq \delta_k \leq 1$ ,  $\lim_{k \rightarrow \infty} \gamma(k, k + \delta_k) = 1/2$ .*

*Proof.* According to Adell and Jodrá (2005, lemma 1),  $\lim_{k \rightarrow \infty} \gamma(k, k) = 1/2$ . From (4),

$$\gamma(k, k + 1) - \gamma(k + 1, k + 1) = (k + 1)^k \exp(-k - 1)/k!. \quad (64)$$

As a result of Lemma 3, the right-hand side of (64) tends to 0 as  $k \rightarrow \infty$ , and therefore  $\lim_{k \rightarrow \infty} \gamma(k, k + 1) = \lim_{k \rightarrow \infty} \gamma(k, k) = 1/2$ . The fact that  $\gamma(t, u)$  is monotone increasing in  $u$  implies that  $\gamma(k, k) \leq \gamma(k, k + \delta_k) \leq \gamma(k, k + 1)$ , and the desired result follows.  $\square$   $\square$

**Lemma 5.** *For  $r \geq 2$  and  $a = b$ , the solution  $\Omega^*$  to (26) lies in  $(r - 4/3, r - 2 + \log 2)$ , and  $\lim_{r \rightarrow \infty} (\Omega^* - r) = -4/3$ .*

*Proof.* The result follows from Alm (2003).  $\square$   $\square$

**Lemma 6.** *For  $r \geq 2$  and  $a = b$ , the solution  $\Omega^*$  to (30) lies in  $(r - 1, r)$ .*

*Proof.* Using (4) the condition (30) can be written, for  $a = b$ , as

$$\left( \frac{\Omega^{*2}}{r(r-1)} + 1 \right) \gamma(r, \Omega^*) = \frac{\Omega^{*r} \exp(-\Omega^*)}{r!} \left( 1 - \frac{\Omega^*}{r-1} \right) + 1. \quad (65)$$

Let  $v_1(\Omega^*)$  and  $v_2(\Omega^*)$  respectively denote the left-hand and right-hand sides of (65), considered as functions of  $\Omega^*$ . It is easily seen that  $v_1$  is monotone increasing, whereas  $v_2$  is monotone decreasing on the interval  $(r - 1, r)$ . From Lemma 5 and the monotonicity of  $\gamma(t, u)$  with respect to  $u$  it follows that  $\gamma(r, r - 1) < 1/2$ , which implies that  $v_1(r - 1) < 1$ . On the other hand,  $v_2(r - 1) = 1$ . Therefore the solution to (65), or equivalently to (30), satisfies  $\Omega^* > r - 1$ . By analogous arguments it is seen that  $v_1(r) > 1$  and  $v_2(r) < 1$ . Therefore the solution satisfies  $\Omega^* < r$ .  $\square$   $\square$

**Lemma 7.** *For any  $\delta_1, \delta_2 \in \mathbb{R}$ , the sequence of functions  $h_k(\delta) = \exp(\delta)(1 + \delta/(k - 1))^{-k+1}$ ,  $k \in \mathbb{N}$ ,  $k \geq 2$ ,  $\delta \in [\delta_1, \delta_2]$  converges uniformly to 1 as  $k \rightarrow \infty$ .*

*Proof.* Let  $k_0 = \max\{-\delta_1, 0\} + 2$ . As  $1 + \delta/(k-1) > 0$  for  $\delta \geq \delta_1, k \geq k_0$ , it is possible to take logarithms in the definition of  $h_k(\delta)$  for  $k \geq k_0$ , which gives  $\log h_k(\delta) = \delta - (k-1) \log(1 + \delta/(k-1))$ . Replacing  $k$  by a continuous variable  $x > 1$  and using the inequality  $\log(1+t) > t/(1+t)$ , it is seen that

$$\frac{\partial}{\partial x} \left[ \delta - (x-1) \log \left( 1 + \frac{\delta}{x-1} \right) \right] = -\log \left( 1 + \frac{\delta}{x-1} \right) + \frac{\delta}{x+\delta-1} < 0. \quad (66)$$

This implies that  $h_{k+1}(\delta) < h_k(\delta)$  for  $k \geq k_0$ . In addition,  $h_k(\delta), \delta \in [\delta_1, \delta_2]$  is a continuous function and converges pointwise to 1 as  $k \rightarrow \infty$ . Thus Dini's theorem (Apostol, 1974, p. 248) can be applied, which ensures that the convergence is uniform.  $\square$

*Proof of Proposition 3.* For  $L$  as in (1), particularizing (24) to  $a = b, \Omega = r-1$  and using (4),

$$\frac{\bar{\eta}}{a} = 2(\gamma(r-1, r-1) - \gamma(r, r-1)) = \frac{2(r-1)^{r-2} \exp(-r+1)}{(r-2)!}. \quad (67)$$

In the following, the value  $\Omega^*$  determined by (26) for a given  $r$  will be denoted as  $\Omega_r^*$ . Particularizing (27) to  $a = b$ ,

$$\frac{\eta^*}{a} = \frac{2\Omega_r^{*r-1} \exp(-\Omega_r^*)}{(r-1)!}. \quad (68)$$

From (67) and (68), with  $h_k(\delta)$  as defined in Lemma 7, it follows that

$$\frac{\bar{\eta}}{\eta^*} = \left( \frac{r-1}{\Omega_r^*} \right)^{r-1} \exp(\Omega_r^* - r + 1) = h_r(\delta_r^*) \quad (69)$$

with  $\delta_r^* = \Omega_r^* - r + 1$ . Lemma 5 establishes that  $\delta_r^* \in [-1/3, -1 + \log 2]$  and  $\lim_{r \rightarrow \infty} \delta_r^* = -1/3$ . On the other hand, by Lemma 7,  $h_k \rightarrow 1$  uniformly on  $[-1/3, -1 + \log 2]$  as  $k \rightarrow \infty$ . Therefore, according to Apostol (1974, theorem 9.16),  $\lim_{k,l \rightarrow \infty} h_k(\delta_l^*)$  exists and equals 1. Thus, in particular,  $\lim_{r \rightarrow \infty} h_r(\delta_r^*) = 1$ , which combined with (69) establishes that  $\lim_{r \rightarrow \infty} \bar{\eta}/\eta^* = 1$ .

For  $L$  as in (2), and with  $\Omega^*$  given by (30), let  $\Omega_r^*$  and  $\delta_r^*$  be defined as before. In addition, let  $\tilde{\Omega}_r$  denote the value of  $\tilde{\Omega}$  corresponding to a given  $r$ , and  $\tilde{\delta}_r = \tilde{\Omega}_r - r + 1$ . Particularizing (28) to  $a = b, \Omega = \tilde{\Omega}_r$  and using (4) gives

$$\begin{aligned} \frac{\bar{\eta}}{a} &= \frac{r}{\tilde{\Omega}_r} (\gamma(r-1, \tilde{\Omega}_r) - \gamma(r+1, \tilde{\Omega}_r)) + \left( \frac{\tilde{\Omega}_r}{r-1} - \frac{r}{\tilde{\Omega}_r} \right) \gamma(r-1, \tilde{\Omega}_r) + \frac{r}{\tilde{\Omega}_r} - 1 \\ &= \frac{\tilde{\Omega}_r^{r-1} \exp(-\tilde{\Omega}_r)}{(r-1)!} \left( 1 + \frac{r}{\tilde{\Omega}_r} \right) + \left( \frac{\tilde{\Omega}_r}{r-1} - \frac{r}{\tilde{\Omega}_r} \right) \gamma(r-1, \tilde{\Omega}_r) + \frac{r}{\tilde{\Omega}_r} - 1. \end{aligned} \quad (70)$$

Thus  $\bar{\eta}$  can be written as  $a(\tilde{\theta}_0 + \tilde{\theta}_1 + \tilde{\theta}_2)$  with

$$\tilde{\theta}_0 = \frac{(r-1 + \tilde{\delta}_r)^{r-2} \exp(-r+1 - \tilde{\delta}_r)}{(r-2)!} \left( 2 + \frac{1 + \tilde{\delta}_r}{r-1} \right), \quad (71)$$

$$\tilde{\theta}_1 = \frac{\tilde{\delta}_r^2}{r-1} + \frac{2\tilde{\delta}_r - 1}{r-1 + \tilde{\delta}_r} \gamma(r-1, r-1 + \tilde{\delta}_r), \quad (72)$$

$$\tilde{\theta}_2 = \frac{1 - \tilde{\delta}_r}{r-1 + \tilde{\delta}_r}. \quad (73)$$



The quotient  $\tilde{\theta}_2/\tilde{\theta}_0$  is computed as

$$\frac{\tilde{\theta}_2}{\tilde{\theta}_0} = \frac{(1 - \tilde{\delta}_r) \exp(\tilde{\delta}_r)}{\left(1 + \frac{\tilde{\delta}_r}{r-1}\right)^{r-1} \left(2 + \frac{1+\tilde{\delta}_r}{r-1}\right)} \cdot \frac{(r-1)! \exp(r-1)}{(r-1)^{r-1/2}} \cdot \frac{1}{\sqrt{r-1}}. \quad (74)$$

Proposition 1 implies that  $\tilde{\delta}_r \in (0, 1)$ . Taking into account that  $r \geq 2$ , it is seen that the first factor in (74) lies in a bounded interval for all  $r$ , whereas, by the equality in Lemma 3, the second factor tends to  $\sqrt{2\pi}$  as  $r \rightarrow \infty$ . As a result,  $\lim_{r \rightarrow \infty} \tilde{\theta}_2/\tilde{\theta}_0 = 0$ . Similarly,  $\tilde{\theta}_1/\tilde{\theta}_0$  is expressed as

$$\frac{\tilde{\theta}_1}{\tilde{\theta}_0} = \frac{\left(\frac{\tilde{\delta}_r^2}{r-1} + 2\tilde{\delta}_r - 1\right) \exp(\tilde{\delta}_r)}{\left(1 + \frac{\tilde{\delta}_r}{r-1}\right)^{r-1} \left(2 + \frac{1+\tilde{\delta}_r}{r-1}\right)} \cdot \frac{(r-1)! \exp(r-1)}{(r-1)^{r-1/2}} \cdot \gamma(r-1, r-1+\tilde{\delta}_r) \cdot \frac{1}{\sqrt{r-1}}. \quad (75)$$

As before, the first factor in the right-hand side of (75) is bounded, and the second tends to  $\sqrt{2\pi}$ . The third factor tends to  $1/2$  by Lemma 4. Thus  $\lim_{r \rightarrow \infty} \tilde{\theta}_1/\tilde{\theta}_0 = 0$ .

The quotient  $\eta^*/a$  is given as in (70) with  $\tilde{\Omega}_r$  replaced by  $\Omega_r^*$ ; and  $\eta^* = a(\theta_0^* + \theta_1^* + \theta_2^*)$ , where  $\theta_0^*$ ,  $\theta_1^*$  and  $\theta_2^*$  are obtained from (71)–(73) with  $\tilde{\delta}_r$  replaced by  $\delta_r^*$ . Lemma 6 implies that  $\delta_r^* \in (0, 1)$ , and arguments analogous to those in the preceding paragraph show that  $\theta_1^*/\theta_0^*$  and  $\theta_2^*/\theta_0^*$  tend to 0 as  $r \rightarrow \infty$ . As a result,  $\lim_{r \rightarrow \infty} \bar{\eta}/\eta^*$  can be computed as

$$\lim_{r \rightarrow \infty} \frac{\bar{\eta}}{\eta^*} = \lim_{r \rightarrow \infty} \frac{\tilde{\theta}_0}{\theta_0^*} = \lim_{r \rightarrow \infty} \frac{\left(1 + \frac{\tilde{\delta}_r}{r-1}\right)^{r-2} \exp(-\tilde{\delta}_r)}{\left(1 + \frac{\delta_r^*}{r-1}\right)^{r-2} \exp(-\delta_r^*)} \cdot \frac{2 + \frac{1+\tilde{\delta}_r}{r-1}}{2 + \frac{1+\delta_r^*}{r-1}}. \quad (76)$$

Since  $\tilde{\delta}_r, \delta_r^* \in (0, 1)$  for all  $r$ , it is clear that the second factor in the rightmost part of (76) tends to 1 as  $r \rightarrow \infty$ . By Lemma 7,  $(1 + \delta/(r-1))^{r-1} \exp(-\delta) \rightarrow 1$  uniformly for  $\delta \in (0, 1)$ . This implies that the numerator and denominator of the first factor in (76) tend to 1 as  $r \rightarrow \infty$  (note that  $\tilde{\delta}_r$  and  $\delta_r^*$  are not required to converge). Consequently  $\lim_{r \rightarrow \infty} \bar{\eta}/\eta^* = 1$ . □ □

## References

- Abramowitz M, Stegun IA (eds) (1970) Handbook of Mathematical Functions, ninth edn. Dover
- Adell JA, Jodrá P (2005) The median of the Poisson distribution. *Metrika* 61:337–346
- Akdeniz F (2004) New biased estimators under the linex loss function. *Statistical Papers* 45:175–190
- Alm SE (2003) Monotonicity of the difference between median and mean of gamma distributions and of a related Ramanujan sequence. *Bernoulli* 9(2):351–371
- Alvo M (1977) Bayesian sequential estimation. *Annals of Statistics* 5(5):955–968
- Apostol TM (1974) *Mathematical Analysis*, 2nd edn. Addison-Wesley
- Baran J, Magiera R (2010) Optimal sequential estimation procedures of a function of a probability of success under LINEX loss. *Statistical Papers* 51(3):511–529

- Berger JO (1985) *Statistical Decision Theory and Bayesian Analysis*, 2nd edn. Springer-Verlag
- Best DJ (1974) The variance of the inverse binomial estimator. *Biometrika* 61(2):385–386
- Cabilio P (1977) Sequential estimation in Bernoulli trials. *Annals of Statistics* 5(2):342–356
- Cabilio P, Robbins H (1975) Sequential estimation of  $p$  with squared relative error loss. *Proceedings of the National Academy of Sciences of the United States of America* 72(1):191–193
- Christoffersen PF, Diebold FX (1997) Optimal prediction under asymmetric loss. *Econometric Theory* 13:808–817
- DeGroot MH (1959) Unbiased sequential estimation for binomial populations. *Annals of Mathematical Statistics* 30(1):80–101
- Girshick MA, Mosteller F, Savage LJ (1946) Unbiased estimates for certain binomial sampling problems with applications. *Annals of Mathematical Statistics* 17(1):13–23
- Granger CWJ (1969) Prediction with a generalized cost of error function. *Operational Research Quarterly* 20(2):199–207
- Haldane JBS (1945) On a method of estimating frequencies. *Biometrika* 33(3):222–225
- Hubert SL, Pyke R (2000) Sequential estimation of functions of  $p$  for Bernoulli trials. In: *Game Theory, Optimal Stopping, Probability and Statistics*, Institute of Mathematical Statistics, pp 263–294
- Lehmann EL, Casella G (1998) *Theory of Point Estimation*, 2nd edn. Springer
- Mendo L (2009) Estimation of a probability with guaranteed normalized mean absolute error. *IEEE Communications Letters* 13(11):817–819
- Mendo L (2012) Asymptotically optimum estimation of a probability in inverse binomial sampling. *Journal of Statistical Planning and Inference* 142(10):2862–2870.
- Mendo L, Hernando JM (2006) A simple sequential stopping rule for Monte Carlo simulation. *IEEE Transactions on Communications* 54(2):231–241
- Mendo L, Hernando JM (2008) Improved sequential stopping rule for Monte Carlo simulation. *IEEE Transactions on Communications* 56(11):1761–1764
- Mendo L, Hernando JM (2010) Estimation of a probability with optimum guaranteed confidence in inverse binomial sampling. *Bernoulli* 16(2):493–513
- Mikulski PW, Smith PJ (1976) A variance bound for unbiased estimation in inverse sampling. *Biometrika* 63(1):216–217
- Sathe YS (1977) Sharper variance bounds for unbiased estimation in inverse sampling. *Biometrika* 64(2):425–426