Sequential Estimation of Functions of p for Bernoulli Trials

Steven L. Hubert and Ronald Pyke
University of Washington

Abstract

We investigate the problem of estimating a function of p, the probability of success in a sequence of independent identically distributed Bernoulli trials. The estimation problems of interest here are those which become more difficult when p is small, such as when estimating the rate, $1/p$, with squared error loss, or when estimating p with relative squared error loss. A sequential estimation procedure is required when the specification of the problem requires that the solution be valid for all values of p. The formulation used is typical of those taken in the literature, but a new approach to the asymptotic analysis of the problem is taken, in which the asymptotics are driven by letting the parameter, p, approach zero. Previous approaches to the study of the asymptotics have always fixed the distributional parameters, such as p in the Bernoulli case, and allowed some other parameter of the problem (e.g., accuracy or cost) to vary. The emphasis in this paper is on the estimation of powers of p with a loss function equal to squared error loss multiplied by another power of p. Consistency and efficiency are defined in this framework and conditions under which the procedure is consistent and efficient are given.

1. Introduction. There are two major reasons for using sequential procedures in inference. The first, and possibly the main reason with regard to *hypothesis testing*, is to decrease expected sample size. The second reason is to improve tractability; an inference problem may be impossible to solve otherwise. An early *estimation* example of the latter is Stein's 1945 two-stage procedure for estimating the mean of a normal distribution with unknown variance, a problem for which no fixed sample size procedure suffices for all possible values of the variance.

AMS 1991 subject classification: 62L12, 62F12.

Key words and phrases: sequential estimation, binomial model, uniform integrability.

Further improvements should be possible by increasing the number of stages, and ultimately, of course, sequential procedures should treat each observation as a "stage"; for the problem of estimating the mean of a normal distribution when the variance is unknown, the approximate solution given by Anscombe (1953) has this structure. This is reviewed in the next section as motivation for the specific problems and procedures of this paper which are summarized as follows.

Let $\{X_i : i \geq 1\}$ be a sequence of independent Bernoulli-p random variables (rv's). The general problem of interest here is that of estimating the value $g(p)$ of a smooth function of p. We restrict our attention to a loss function which is a product of squared error loss and a function of p, namely,

$$L_n(t_n, p) = h(p)(t_n - g(p))^2, \tag{1.1}$$

in which t_n is an estimator for $g(p)$. We further restrict ourselves to estimators of the form $t_n = g(\overline{X}_n)$ where $\overline{X}_n$ is the sample mean. Our goal is to find a stopping time N for which the *risk*, $E(L_N)$, is acceptably close to a pre-specified constant c.

The motivation for our choice of stopping time is that if $\overline{X}_n$ is close to p and g is sufficiently smooth, then

$$g(\overline{X}_n) \approx g(p) + g'(p)(\overline{X}_n - p), \tag{1.2}$$

so that

$$\operatorname{Var}(g(\overline{X}_n)) \approx \frac{(g'(p))^2 p}{n} \tag{1.3}$$

for small p. Assuming this approximation is a good one, the best fixed sample size for a given p, say ξ_p, is obtained by solving for n in the approximate equation, risk $\approx h(p)(g'(p))^2 p/n = c$ which yields (neglecting for the moment that an integer is required):

$$\xi_p \approx h(p) g'(p)^2 p/c \tag{1.4}$$

for p small. The stopping time, N is then chosen to be the smallest n for which $n \geq \xi_{\overline{X}_n}$.

In this paper, we restrict ourselves further to the case in which both $g(p)$ and $h(p)$ are powers of p, namely,

$$g(p) = p^a \text{ and } h(p) = p^b, \tag{1.5}$$

for some real a and b. Examples covered by this include estimation of the rate $1/p$ with squared error loss (that is, $a = -1$, $b = 0$ so that $2a + b = -2$) or estimation of p with relative squared error loss (that is $a = 1$, $b = -2$

so that $2a + b = 0$). Under (1.5), $L_n(t_n, p) = p^b(t_n - p^a)^2$ and we take $\xi_p = (a^2/c)p^{2a+b-1}$. Notice first that the sign of the quantity $2a+b$ will play an important role. In particular,, if one takes t_n to be identically zero for every n then $L_n = p^{2a+b}$. Thus if $2a + b > 0$, any stopping rule including $N \equiv 0$, satisfies

$$E(L_N) = p^{2a+b} \to 0 \tag{1.6}$$

as $p \to 0$. This shows that if $2a + b$ is positive, one can estimate p^a with arbitrarily small risk as $p \to 0$ *without taking any observations*! To avoid these trivial cases, we assume that $2a + b \leq 0$. Situations with $2a + b = 0$ are boundary cases that may well be of interest as discussed briefly in the following.

If $2a + b = 0$, then $\xi_p = a^2/cp$. Hence the stopping time becomes

$$\begin{aligned} N &= \inf\{n > 0 : n \geq \xi_{\overline{X}_n}\} = \inf\{n > 0 : n\overline{X}_n \geq a^2/c\} \\ &= \inf\{n > 0 : S_n \geq a^2/c\} \end{aligned} \tag{1.7}$$

where $S_n = X_1 + \cdots + X_n$. Thus N is just the waiting time until the $\lceil a^2/c \rceil$'th success, where $\lceil x \rceil$ stands for the smallest integer $\geq x$.

The particular case $a = 1$, $b = -2$ for which $2a + b = 0$ is precisely the case examined by Robbins and Siegmund (1974). They ask for the risk to be equal to a constant c, the same as here, but their asymptotic study of the problem takes place as this constant is allowed to approach zero for fixed p (whereas we fix the constant c and allow p to approach zero). For this situation they obtain the following efficiency and consistency results;

$$\frac{N}{\xi_p} \xrightarrow{P} 1 \text{ as } c \to 0 \tag{1.8}$$

and

$$\frac{\text{risk}}{c} \to 1 \text{ as } c \to 0. \tag{1.9}$$

Cabilio and Robbins (1975) also examine this case but they add a constant cost per observation to the loss function. They then show that (1.8) and (1.9) continue to hold for any fixed p and fixed c, but now it is the cost per observation rather than c that goes to zero. The convergence, however, is not uniform in p in this case, and they note that the procedure does not perform well as $p \to 0$. They introduce a uniform prior distribution on p to address this latter shortcoming.

Consider what happens in the above cases of $2a+b = 0$ if we fix the level of risk c and allow p to converge to zero. Since as shown in (1.7), N is the waiting time until $\lceil a^2/c \rceil$ Bernoulli successes, in order that $\frac{N}{\xi_p} \xrightarrow{P} 1$ as $p \to 0$

it must be true that for every $\varepsilon > 0$, $P(N < (1-\varepsilon)\xi_p)$ converges to zero as $p \to 0$. But $P(N < (1-\varepsilon)\xi_p) = P(S_{\lceil(1-\varepsilon)\xi_p\rceil} \geq \lceil a^2/c \rceil)$ where $S_{\lceil(1-\varepsilon)\xi_p\rceil}$ is just the sum of $\lceil(1-\varepsilon)\xi_p\rceil$ independent Bernoulli trials, so it has a binomial distribution. When $2a + b = 0$, $(1-\varepsilon)\xi_p p = (1-\varepsilon)a^2/c$, a constant, so that the distribution of $S_{\lceil(1-\varepsilon)\xi_p\rceil}$ converges to a Poisson distribution as $p \to 0$. Consequently, the probability that N is less than $(1-\varepsilon)\xi_p$ is nearly constant as $p \to 0$. It cannot, therefore, be true that $\frac{N}{\xi_p} \xrightarrow{P} 1$ as $p \to 0$ when $2a+b=0$. In any event, it is not clear that the above estimation procedure that uses the stopping time N defined in (1.7) and then estimates p^a with $\overline{X}^a$ is a good procedure for small p when $2a + b = 0$.

We assume throughout the rest of this paper that

$$2a + b < 0. \tag{1.10}$$

Section 2 is a review of related literature. Section 3 gives a proof of the efficiency of the stopping time in the sense that

$$\frac{N}{\xi_p} \xrightarrow{P} 1 \text{ as } p \to 0. \tag{1.11}$$

Section 4 is a discussion of the asymptotic distribution of N and of the estimator $g(\overline{X}_N)$. Section 5 gives the proof of uniform integrability for the stopping times and loss functions. It is this uniform integrability together with the earlier convergence properties that enables one to interchange integrations and limits to obtain the key *consistency* result:

$$\text{risk} = E(L_n) \to c \text{ as } p \to 0. \tag{1.12}$$

We conclude in Section 6 with some remarks and open questions.

2. A Review of Sequential Estimation. Sequential estimation for Bernoulli trials is closely related to sequential estimation in general. The method of deciding when to stop sampling is almost always the same and is fairly obvious, as are the estimators that are usually proposed. However, the mathematical analyses of the various problems differ in the goal pursued and in their difficulty. The following is a brief historical review. The references cited below are but a few of the many papers on sequential estimation, though they are the ones pertaining most directly to the topic of this paper.

The first problem considered was the estimation of the mean of a normal distribution when the variance is unknown. Since there is no fixed sample size procedure which can work well for all values of the variance, σ^2, a sequential procedure must be used. The usual estimation procedure is to estimate the mean, μ, with the sample mean, $\overline{X}_N$, once you stop. Typically, the approach is to notice that, if σ were known, there would be a number

$n(\sigma)$ that gives the desired accuracy. That is, for any $n \geq n(\sigma)$, $\overline{X}_n$ would be precise enough for the problem at hand. Since σ is unknown it is estimated after n observations, by $\widehat{\sigma}_n$, say. Then, the stopping rule will be to stop the first time, N say, that $n \geq n(\widehat{\sigma}_n)$ and at that time estimate μ with the sample mean, $\overline{X}_N$,

Two basic goals have been pursued in the literature:

1. Find an interval $(\overline{X}_N - d, \overline{X}_N + d)$ which contains μ with *confidence* $\geq \gamma$. Subject to this minimize $E(N)$, the expected stopping time.

2. Define a loss function like $L_n = A(\overline{X}_n - \mu)^2 + cn$ where c is the cost per observation. Find a stopping rule, N, which minimizes the risk = $E(L_N)$.

Stein (1945) proposed a two-stage procedure which gave a confidence interval for μ but his procedure entails a high value for $E(N)$ when the choice of the initial sample size is bad. His procedure does, however, give a guaranteed *confidence* $\geq \gamma$ whereas many of the later procedures only give *confidence* $\rightarrow \gamma$. One can certainly expect to be able to utilize the information about σ more efficiently than is allowed by a two-stage procedure.

Wald (1951) tried a slightly different approach to a similar problem: He restricted his attention to one parameter families with densities and found procedures which are asymptotically minimax. He looked at the maximum risk over all possible values of the parameter and found the rule that minimizes it, at least when the cost per observation approaches zero. His rule is basically to stop the first time that the estimated decrease in variance of the estimator is less than the cost per observation. This rule is the same as a rule proposed by Starr and Woodroofe (1972), to be discussed later, which is shown to have some nice properties in the exponential case. Wald's proof of asymptotic minimaxity does not apply to the exponential distribution since he requires the assumption that the information is bounded away from zero, which is not true in that case.

Anscombe (1953) found an approximate solution for the basic normal problem, namely; find a confidence interval for the mean of a normal distribution having width $2d$ and confidence coefficient γ. The procedure should work for all values of the variance σ^2. It is also desirable to have the random sample size, N, as small as possible. This is just a restatement of goal (1) above for the normal case. Anscombe's solution is a very natural one. Let $\alpha = \Phi^{-1}(1 - \gamma/2)$ be the appropriate percentile of the standard normal distribution. Then, if σ were known, the required sample size would be $n(\sigma) = \alpha^2\sigma^2/d^2$. Let N be the smallest value of n larger than $n(\widehat{\sigma}_n)$, where $\widehat{\sigma}_n^2$ is the usual unbiased estimate of σ^2 after n observations have been made.

Then, there are two conclusions:

$$P(\text{coverage}) \to \gamma \text{ as } d \to 0 \tag{2.1}$$

and

$$\frac{E(N)}{n(\sigma)} \to 1 \text{ as } d \to 0, \tag{2.2}$$

where P and E refer to the true underlying $N(\mu, \sigma^2)$ distribution. Chow and Robbins (1965) extended this result to all distributions of finite variance.

Expression (2.1) provides a type of *consistency*, whereas, expression (2.2) represents a type of *efficiency*, since it compares the expected sample size of the procedure with the optimal fixed sample sizes. Instead of the asymptotic consistency property in (2.1), it would be even better to have guaranteed confidence for all values of d and σ, but there are only limited results in this direction. In the normal case, the desired result would be

$$P(\text{coverage}) \geq \gamma \text{ for all } d \text{ and } \sigma,$$

rather than (2.1). Toward this end, Starr (1966a, 1966b) computed a lower bound on the confidence coefficient for all values of σ/d for the case of the normal distribution. He showed, for example, that the actual confidence is greater than or equal to .928 for all d and σ when the nominal value of γ is .95. Starr also approached the problem using the second goal mentioned above concerning the loss plus cost per observation criterion. He showed that it is important to have a minimum allowable sample size. In the following, let L_n be the loss function and suppose m is the minimum allowable sample size so that the restriction $N \geq m$ is imposed on the stopping times considered. (A similar minimum sample size restriction will also be important for the uniform integrability results of this paper.) For some $s > 0$, consider

$$L_n = A|\overline{X}_n - \mu|^s + n, \text{ and } N = \inf\{n \geq m : n \geq K\widehat{\sigma}_n^{\frac{2s}{s+2}}\}.$$

This procedure is efficient if the minimum sample size m is greater than $s^2/(s+2) + 1$. So, for squared error loss, you need $m \geq 3$ to have an efficient procedure. Starr adapted this approach to the fixed-width confidence interval problem and obtained all moments of N.

One can hope for a stronger result than (2.2). One such result might be that the *difference* between $E(N)$ and $n(\sigma)$, called the *regret*, is bounded for all values of d and σ. In unpublished work referred to in Simons (1968), Simons and Chernoff strengthened the Chow and Robbins (1965) efficiency result to

$$\limsup_{d \to 0} E(N - n(\sigma)) \leq 1 + m$$

(instead of $N/n(\sigma) \to 1$). Simons' main result, though, is that you can achieve guaranteed confidence by taking a *fixed* number of observations more than the Anscombe rule says to take. This number, k say, depends only on γ. He also obtained the result that his procedure has bounded regret if a minimum example size of three is used, that is,

$$E((N+k) - n(\sigma)) \leq m + k.$$

Starr and Woodroofe (1968, 1969, 1972) used the loss plus cost per observation approach to get a procedure with *regret* ≤ 1 as $d \to 0$ and *confidence* $\geq \gamma - 1.72(1-\gamma)$ uniformly for all d, provided the sample size is sufficiently large. They showed for $L_n = A|\overline{X}_n - \mu|^s + n$, that the procedure has bounded excess risk (excess over the best fixed sample size procedure for known σ) if and only if the minimum sample size $m \geq s+1$.

Woodroofe (1977) sharpened some of the results for the Gaussian problem by using more precise approximations. His results apply for a sufficiently large sample size (either d sufficiently close to zero or σ sufficiently large). He obtains a guaranteed confidence procedure by taking k extra observations for $k > 2 - \nu$, where ν is a computable constant. So, if the sample size is large, only two extra observations are needed to obtain $P(\text{coverage}) \geq \gamma$. This result is independent of the value of γ but does require a minimum sample size $m \geq 7$. It also requires a sufficiently small d, whereas, Simons' (1968) result is true for all d. Of course, Simons does not say how many more observations are needed, only that there is some finite k which will work.

The above fixed-width confidence interval problem can be formulated in another essentially equivalent way. That is, specify a loss function such as squared error loss, and ask that the expected loss be equal to some prescribed constant. Asymptotic analysis can then be done by letting the prescribed constant approach zero. This analysis is very similar to the analysis done when letting d approach zero in (2.1) and (2.2). A further formulation of asymptotics is the following: Let the underlying distribution converge to a degenerate one. For example, in the Gaussian case this could mean letting the variance of the distribution go to infinity, making it more and more difficult to get a fixed-width confidence interval.

A Bayesian approach to sequential estimation was used in Alvo (1977). Instead of looking at the worst possible parameter value θ or all θ's, the procedure is to integrate the risk with respect to a prior distribution on θ. Alvo defined *regret* as

(Bayes risk of the procedure) − (Bayes risk of the optimal Bayes procedure).

He used an information inequality approach to get a lower bound on the Bayes risk of the optimal procedure and then found rules with *regret* bounded by terms of order c (cost per observation) for several one parameter exponential families.

Vardi (1979a) used Alvo's theorem to give some justification for the bounded *regret* approach of Robbins (1959), Starr (1966a, 1966b), Starr and Woodroofe (1968, 1969, 1972), Woodroofe (1977), and Simons (1968). If a procedure has bounded *regret* then it is asymptotically Bayes with respect to any sequence of priors which smooth out towards infinity in a suitable fashion. He also showed that no procedure can have uniformly negative *regret*. Both of these results justify the search for bounded *regret* procedures. Vardi (1979b) obtained such a procedure for the Poisson case.

The Gaussian problem discussed here turns out to be one of the simplest applications of sequential estimation techniques, but it illustrates the principles well. There are also some results specifically for the exponential (Starr and Woodroofe (1972), and Woodroofe (1982)) and Poisson (Vardi (1979b)) cases, and several for the Bernoulli case (Alvo (1977), Cabilio (1977), Cabilio and Robbins (1975), Haldane (1945, 1956), Robbins and Siegmund (1974), and Siegmund (1982)). In these cases the stopping variable N and the estimator are dependent, whereas in the Gaussian case they are actually independent. This dependence makes the analyses much more difficult.

The focus of this paper is upon the Bernoulli case. There are times when one wishes to estimate p or a function of p more accurately when p is small. For example, one might be using relative error or squared relative error in the loss function, or one may wish to estimate a rate $1/p$. If it were known that p lay between 0.1 and 0.9, say, then a fixed sample size procedure could be used. But if no lower and upper bounds on p can be predetermined, no fixed sample size procedure will suffice.

There has been considerable work on sequential solutions to this problem using an approach similar to that used by Anscombe in the normal problem and leading to analogues of his results. The problem is to find a fixed-width confidence interval for some function of p having a fixed width $2d$ or an expected loss equal to some prescribed constant. The asymptotic analysis of this problem has usually been studied when the fixed width or the prescribed constant is allowed to approach zero. The problem has also been analyzed when the loss function includes some cost per observation and then that cost is allowed to approach zero.

The usual fixed sample size estimate of p has standard error proportional to $\sqrt{p}$. Haldane (1945) notes that " ... it may be desired that the standard error at each value of p should be roughly proportional to p rather than to its square root." This, of course, implies that a more accurate estimate should be used when p is small so that the coefficient of variation is approximately constant. Inverse binomial sampling achieves this. Let N be the number of observations until c successes have been observed. Haldane showed that

$\widehat{p} = \frac{c-1}{N-1}$ is unbiased for p and has standard error approximately proportional to p for small values of p.

Negative binomial sampling illustrates well the motivation behind the stopping procedures of all sequential Bernoulli estimators. It works since a smaller p will, on average, force a larger sample size. Robbins and Siegmund (1974) give a similar example. They wish to estimate p with expected squared relative error equal to some preassigned constant, namely $E(\widehat{p}-p)^2/p^2q^2 \approx c$. For any fixed sample size n the expected loss is $1/npq$. In order for the expected loss to be close to c then they need $npq \approx 1/c$. Since p is unknown, the estimate $\overline{X}_n = S_n/n$ is substituted for p. Since they want to have $n\overline{X}_n(1-\overline{X}_n) \approx 1/c$, they define N by

$$N = \inf\{n > 0 : S_n(n - S_n) > \frac{n}{c}\}.$$

After stopping, $\overline{X}_N$ is used to estimate p. They show that

$$\frac{E(\overline{X}_N - p)^2}{cp^2q^2} \to 1 \text{ as } c \to 0$$

for any fixed $0 < p < 1$ so that as more and more accuracy is required of the estimator their goal of having the expected loss approximately equal to c is achieved.

Cabilio and Robbins (1975) consider the second goal, using the same loss function as above but with a cost per observation added on. They first find a procedure which works well for any fixed p as the unit cost goes to zero, which is similar to Robbins and Siegmund (1974) above. They notice that this procedure performs poorly for fixed cost as $p \to 0$, a problem they avoid by introducing a uniform prior distribution on p. The Bayes procedure for the uniform prior distribution also works well as the cost approaches zero.

Siegmund (1982) strengthens the result of Robbins and Siegmund (1974) by showing that the results obtained there actually hold uniformly in p. They had originally just stated that the expected loss converged as $c \to 0$ for any fixed $0 < p < 1$.

Instead of letting the cost per observation or the width of the confidence interval go to zero, it is very reasonable to *let the parameter p approach zero*. After all, one is not really interested in what happens when one has a minuscule confidence interval or when the cost of observations is tiny, but one *is* interested in what happens when p is small for fixed confidence or fixed cost. This new approach is closely analogous to the usual asymptotic theory for fixed sample size procedures. For these, one is interested in an approximation for large values of n, so one considers a sequence of sample sizes going to infinity and embeds the large n of interest in that sequence. Here, one is interested in an approximation for small p, so one embeds the small p in a sequence of p's going to zero. This is the approach taken in this paper.

3. The Efficiency of the Stopping Time. Here is a formal description of the problem and the estimation procedure to be studied. Let S_n be the partial sum of n independent Bernoulli random variables with success probability p.

Given constants a, b, and c with $a \neq 0$, $2a + b < 0$, and $c > 0$, we wish to estimate the power of p, $g(p) = p^a$. We would like the *risk* to be close to the specified constant c when the loss function is weighted squared error,

$$L_n(t_n, p) = h(p)(t_n - g(p))^2, \tag{3.1}$$

and the estimator $t_n = g(\overline{X}_n)$ where $\overline{X}_n = S_n/n$, and $h(p) = p^b$. As we saw in (1.4) the approximate ideal fixed-sample size for a specific (but unknown) p is given by

$$\xi_p \equiv (a^2/c)p^{2a+b-1} \equiv (A/p)^{1/(1-\rho)} \tag{3.2}$$

where

$$A = (a^2/c)^{1-\rho} \text{ and } 0 < \rho \equiv \frac{2a+b}{2a+b-1} < 1. \tag{3.3}$$

The parameter ρ plays a central role in this paper. Let the stopping time N be defined by

$$N = \inf\{n > 0 : n \geq \xi_{\overline{X}_n}\}. \tag{3.4}$$

Theorem 3.1. (Efficiency) *If* $2a + b < 0$ *then* $\frac{N}{\xi_p} \xrightarrow{P} 1$ *as* $p \to 0$.

Proof. First, simplify the definition of N by rewriting the inequality in (3.4) in terms of S_n and n, namely,

$$\begin{aligned} N &= \inf\{n > 0 : n \geq \xi_{\overline{X}_n}\} \\ &= \inf\{n > 0 : n \geq (a^2/c)(S_n/n)^{-1/(1-\rho)}\} \\ &= \inf\{n > 0 : S_n \geq C(n)\} \end{aligned} \tag{3.5}$$

where

$$C(n) \equiv An^\rho. \tag{3.6}$$

The stopping rule then is to stop as soon as the random walk, S_n, exceeds the curve, $C(n)$. Since $\rho > 0$ the curve is increasing and concave. *Note that* ξ_p *is the solution in* x *of* $C(x) = xp$, *the place where the expectation* *line crosses the curve C.* To prove the theorem, we must show that for any given $\varepsilon > 0$, S_n crosses the curve for the first time between $(1-\varepsilon)\xi_p$ and $(1+\varepsilon)\xi_p$ with probability tending to one as $p \to 0$. We break the proof into two lemmas. Lemma 3.2 states that the stopping time is not likely to be very small, while Lemma 3.3 proves that the stopping time is between the two endpoints above with high probability, given that the random walk has not stopped early.

Let M_p be the value of x where the difference $C(x) - xp$ is at its maximum. In other words, M_p is the point where the tangent to the curve C equals p. Since $C'(x) = A\rho x^{-(1-\rho)}$, then

$$M_p = (\rho A/p)^{1/(1-\rho)}. \tag{3.7}$$

Lemma 3.2. *$P(N < M_p) \to 0$ as $p \to 0$.*

Proof. By definition,

$$P(N < M_p) = P\left(\sup_{1 \le j \le M_p} \frac{S_j - jp}{C(j) - jp} \ge 1\right). \tag{3.8}$$

The denominator, $C(j) - jp$, is an increasing sequence of positive constants on the range $j = 1, 2, \cdots, M_p$. This is clear from the definition of M_p and also from the following since the derivative of $C(x) - xp$ is non-negative if and only if $A\rho\frac{1}{p} \ge x^{1-\rho}$ which is equivalent to $x \le M_p$. Therefore, the Hajek-Rényi (1955) inequality is applicable and it yields

$$P\left(\sup_{1 \le j \le M_p} \frac{S_j - jp}{C(j) - jp} \ge 1\right) \le \sum_{j=1}^{M_p} \frac{p}{(C(j) - jp)^2}. \tag{3.9}$$

Let

$$h_p(x) = \frac{p}{(C(x) - xp)^2}$$

and notice that $h_p(x)$ is decreasing in x for $0 \le x \le M_p$. Combining (3.8) and (3.9) we have

$$P(N < M_p) < \sum_{j=1}^{M_p} h_p(j) \le \int_0^{M_p} h_p(x)dx \equiv H(p) .$$

Introduce $g_p(x) = h_p(x) \cdot 1_{(0,M_p](x)}$ so that

$$H(p) = \int_0^{\infty} g_p(x)dx .$$

Now for fixed $x > 0$ the function $h_p(x)$ is eventually decreasing as $p \to 0$. In particular, it is decreasing for all $p \leq p_0$ where $M_{p_0} \geq x$. For every $x > 0$ there is a largest value of p, p_x say, such that $g_{p_x}(x) > 0$. Clearly, p_x solves $M_{p_x} = x$, so that $p_x = A\rho x^{-(1-\rho)}$.

Since $h_p(x)$ decreases as $p \to 0$ we have

$$\begin{aligned} g(x) &\equiv \sup_{0<p<1} g_p(x) = g_{p_x}(x) = h_{p_x}(x)1_{(0,M_{p_x}]}(x) \\ &= h_{p_x}(x)1_{(0,x]}(x) = h_{p_x}(x) = \frac{p_x}{(C(p_x) - xp_x)^2} \\ &= \frac{A\rho x^{-(1-\rho)}}{\{Ax_\rho - xA\rho x^{-(1-\rho)}\}^2} = A\frac{\rho}{(1-\rho)^2}x^{-(1+\rho)}. \end{aligned}$$

By its definition, g dominates g_p, and, since $\rho > 0$, it is integrable. By dominated convergence and (3.9) we have

$$\lim_{p\to 0} P(N < M_p) \leq \lim_{p\to 0} H(p) = \lim_{p\to 0} \int_0^\infty g_p(x)dx = 0. \qquad \square$$

We now consider the channel determined by two lines of slope p, one on each side of the line $y = xp$ and equidistant from that line. The width of the channel is set so that the upper edge intersects the curve C at the point $(1-\varepsilon)\xi_p$. Since C is concave, the lower edge of the channel will intersect the curve at a point less than $(1+\varepsilon)\xi_p$. If we can show that the random walk, S_n, stays within this channel until it is past the point $(1+\varepsilon)\xi_p$, then we will have shown that N is between $(1-\varepsilon)\xi_p$ and $(1+\varepsilon)\xi_p$. Note first that this channel has vertical width of

$$\begin{aligned} 2\{C((1-\varepsilon)\xi_p) - (1-\varepsilon)\xi_p p\} &= 2\{A(1-\varepsilon)^\rho \xi_p^\rho - (1-\varepsilon)\xi_p p\} \\ &= 2A^{1/(1-\rho)}\widetilde{c}_\varepsilon p^{-\rho/(1-\rho)} \end{aligned} \qquad (3.10)$$

where $\widetilde{c}_\varepsilon = (1-\varepsilon)^\rho - (1-\varepsilon)$.

Lemma 3.3. *For any $\varepsilon > 0$, as $p \to 0$*

$$P(|S_k - kp| \leq (A^{1/(1-\rho)})\widetilde{c}_\varepsilon p^{-\rho/(1-\rho)}; k = 1, 2, \cdots (1+\varepsilon)\xi_p) \to 1.$$

Proof. Use the Kolmogorov inequality in the following form:

$$P(|S_k - k\mu| \leq (\sigma/\delta)\sqrt{L}; 1 \leq k \leq L) \geq 1 - \delta^2\ .$$

in which $\mu = p$, $\sigma = \sqrt{pq}$, $L = (1+\varepsilon)\xi_p$, and

$$\delta = \frac{(pq(1+\varepsilon)\xi_p)^{1/2}}{A^{1/(1-\rho)}\widetilde{c}_\varepsilon} p^{\rho/(1-\rho)}.$$

For any fixed $\varepsilon > 0$ we have $p \to 0 \Rightarrow \delta^2 \to 0$ which proves the lemma. $\square$

We now combine Lemmas 3.2 and 3.3 to complete the proof of Theorem 3.1. Simply observe that

$$\begin{aligned}
& P(|N/\xi_p - 1| \le \varepsilon) \\
= \;& P((1-\varepsilon)\xi_p \le N \le (1+\varepsilon)\xi_p) \\
= \;& P(S_n \text{ stays in the channel and } N \ge M_p) \\
= \;& P(|S_k - kp| \le (a^2/c)\widetilde{c}_\varepsilon p^{-\rho/(1-\rho)}; 1 \le k \le (1+\varepsilon)\xi_p \text{ and } N \ge M_p) \\
= \;& 1 - P(|S_k - kp| > \widetilde{c}_\varepsilon p \xi_p \text{ for some } 1 \le k \le (1+\varepsilon)\xi_p \text{ or } N < M_p) \\
\le \;& 1 - P(|S_k - kp| > \widetilde{c}_\varepsilon p \xi_p \text{ for some } 1 \le k \le (1+\varepsilon)\xi_p) - P(N < M_p)
\end{aligned}$$

and apply the lemmas. □

4. The Asymptotic Normality of the Estimator. The purpose of this section is to establish the asymptotic distribution of the estimator $g(\overline{X}_n) = \overline{X}^a_N$. This is stated below as Corollary 4.1, after we establish the asymptotic normality of N in

Theorem 4.1. $\dfrac{N-\xi_p}{s_p} \xrightarrow{L} N(0,1)$ *as* $p \to 0$, *where*

$$s_p^2 = (1-\rho)^{-2} p^{-1} \xi_p. \tag{4.1}$$

Proof. Observe first of all that if for each $p \in (0,1)$, n_p is a positive integer satisfying $pn_p \to \infty$ as $p \to 0$, then

$$Y_p \equiv \frac{S_{n_p} - pn_p}{\sqrt{pn_p}} \xrightarrow{L} N(0,1) \text{ as } p \to 0. \tag{4.2}$$

This is an easy consequence of Lindeberg's CLT, for example, since $\mathrm{Var}(S_{n_p}) = pn_p(1-p) \to +\infty$ as $p \to 0$. It remains to show that the result to be proved is closely tied to this version of the CLT for Binomials in which the asymptotics are driven by the parameter p approaching zero.

Let $\{\varepsilon_p; p > 0\}$ be a sequence of positive numbers satisfying: $\varepsilon_p \to 0$ as $p \to 0$, $\varepsilon_p > Ks_p/\xi_p$ for a given positive number K, and $(1-\varepsilon_p)\xi_p$ is an integer. Such a choice of ε_p's is possible because

$$s_p/\xi_p = \text{const.} \times p^{\rho/2(1-\rho)} \to 0 \text{ as } p \to 0\ .$$

Since

$$n_p \equiv (1-\varepsilon_p)\xi_p = (1-\varepsilon_p)(A/p)^{1/(1-\rho)},$$

is of the form covered by (4.2), we have

$$Y_p \equiv \frac{S_{(1-\varepsilon_p)\xi_p} - (1-\varepsilon_p)\xi_p p}{((1-\varepsilon_p)\xi_p p)^{1/2}} \xrightarrow{L} N(0,1)$$

as $p \to 0$. Define

$$z_p(x) \equiv \xi_p + x s_p, \tag{4.3}$$

as a quantity that is approximately x standard deviations (of N) away from ξ_p, the approximate mean of N.

Consider the random channel of vertical width $2\varepsilon(pn_p)^{1/2}$ which runs from time $n_p = (1-\varepsilon_p)\xi_p$ to time $z_p(x)$, has slope p, and is centered around the point S_{n_p}. Assume $x > 0$; the case of $x < 0$ will be similar. Let $A_{x,\varepsilon}$ be the event that the random walk S_n stays in this random channel from time $n_p = (1-\varepsilon_p)\xi_p$ to time $z_p(x)$. It is clear that $A_{x,\varepsilon}$ is independent of the past up to time $(1-\varepsilon_p)\xi_p$ so

$$P(A_{x,\varepsilon}) = P(|S_k - kp| \le \varepsilon(pn_p)^{1/2}; k = 1, 2, \cdots, z_p(x) - n_p\xi_p) .$$

Lemma 4.2. *Given any x and $\varepsilon > 0$, $P(A_{x,\varepsilon}) \to 1$ as $p \to 0$.*

Proof. We have

$$\begin{aligned} P(A^c_{x,\varepsilon}) &= P(|S_k - kp| > \varepsilon(pn_p)^{1/2}; \text{some } k = 1, 2, \cdots, z_p(x) - n_p \\ &< \frac{(z_p(x) - n_p pq}{\varepsilon^2 n_p p} \end{aligned} \tag{4.4}$$

by Kolmogorov's Inequality. The right hand side of (4.4) is less than

$$\frac{z_p(x) - (1-\varepsilon_p)\xi_p}{\varepsilon^2(1-\varepsilon_p)\xi_p} = \frac{1}{\varepsilon^2(1-\varepsilon_p)}\left(\varepsilon_p + \frac{xs_p}{\xi_p}\right) \to 0 \text{ as } p \to 0,$$

by (4.2). □

The next result gives a lower bound on the probability of interest.

Lemma 4.3. *For all x*

$$\liminf_{p\to 0} P(N \le z_p(x)) \ge \Phi(x).$$

Proof. For any arbitrary $\varepsilon > 0$,

$$P(N \le z_p(x)) \ge P((N \le z_p(x)) \cap A_{x,\varepsilon}).$$

Let L denote the line (depending upon x) with slope p which intersects the curve C at the point $z_p(x)$, and let $\gamma_{\varepsilon_p}(x)$ be the height of L at the point

$n_p = (1 - \varepsilon_p)\xi_p$. If the event $A_{x,\varepsilon}$ occurs and S_{n_p} is large enough, then N will be less than $z_p(x)$. In other words,

$$\begin{aligned} P(N \leq z_p(x)) \cap A_{x,\varepsilon}) &\geq P(S_{n_p} \geq \gamma_{\varepsilon_p}(x) + \varepsilon(pn_p)^{1/2}) \cap A_{x,\varepsilon}) \quad (4.5) \\ &= P(S_{n_p} \geq \gamma_{\varepsilon_p}(x) + \varepsilon(pn_p)^{1/2}) P(A_{x,\varepsilon}) \end{aligned}$$

by independence. The function γ_{ε_p} is complicated, so we define a simpler function, $\delta_{1,\varepsilon_p}(.)$, as follows that will satisfy $\delta_{1,\varepsilon_p} \geq \gamma_{\varepsilon_p}$: Let L_1 be the line tangent to the curve C at the point ξ_p. Let L_2 be the line with slope p which intersects L_1 at the point $z_p(x)$. Then define $\delta_{1,\varepsilon_p}(x)$ to be the height of L_2 at the point $(1-\varepsilon_p)\xi_p$. See Figure 4.1 for a sketch of these lines when $x > 0$; when $x < 0$ one would have $z_p(x) < \xi_p$ and a corresponding relocation of the lines relative to the dotted expectation line.

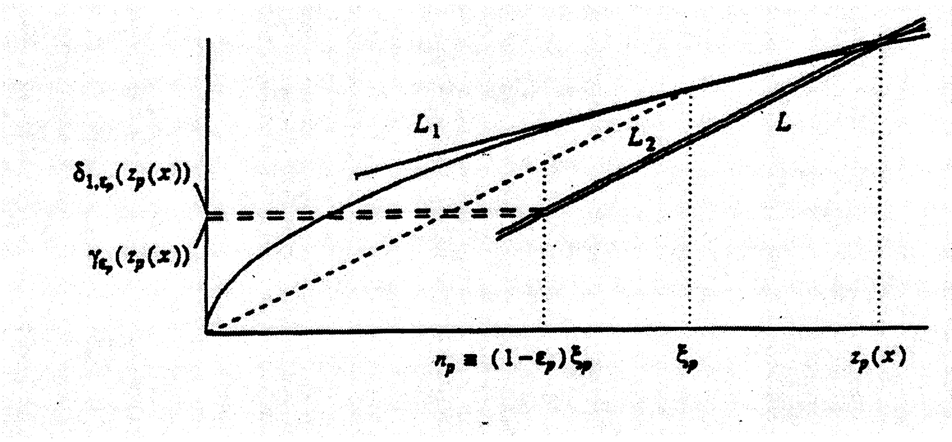

Figure 4.1

We require the following technical result about δ_{1,ε_p}.

Proposition 4.4. *Given $\eta > 0$ there exists a $p_0 > 0$ such that*

$$\frac{\delta_{1,\varepsilon_p}(x) - pn_p}{\sqrt{pn_p}} < -x + \eta$$

for all $p < p_0$ and for all x.

Proof. For horizontal and vertical variables w and $y(w)$, the equation for the line L_1 is

$$y(w) = C(\xi_p) + C'(\xi_p)(w - \xi_p). \tag{4.6}$$

Now, $C'(x) = A\rho x^{-(1-\rho)}$ so that direct substitution shows that $C'(\xi_p) = \rho p$. It then follows that the equation for L_1 is

$$y(w) = C(\xi_p) + \rho p(w - \xi_p),$$

and for L_2 it is

$$y(w) = pw - pn_p + \delta_{1,\varepsilon_p}(x).$$

Since L_1 and L_2 intersect at $z_p(x)$, setting the two equations equal there yields

$$\delta_{1,\varepsilon_p}(x) - pn_p = C(\xi_p) + \rho p(z_p(x) - \xi_p) - pz_p(x). \tag{4.7}$$

Since $z_p(x) - \xi_p = xs_p$ and $C(\xi_p) = p\xi_p$ one obtains from (4.7) that

$$\begin{aligned}\delta_{1,\varepsilon_p}(x) - pn_p p &= p\xi_p - pz_p(x) + \rho p(z_p(x) - \xi_p) \\ &= -(1-\rho)pxs_p = -x\xi_p^{1/2}p^{1/2}.\end{aligned}$$

But this means that $\delta_{1,\varepsilon_p}(x) - pn_p = -x(1-\varepsilon_p)^{-1/2}\sqrt{pn_p}$. The proof of the proposition is therefore complete since $\varepsilon_p > 0$ and $\varepsilon_p \to 0$ as $p \to 0$. □

Return now to (4.5) and the proof of Lemma 4.4. Since $\delta_{1,\varepsilon_p} \geq \gamma_{\varepsilon_p}$, (4.5) gives

$$\begin{aligned}P(N \leq z_p(x)) \geq P(S_{n_p} &\geq \delta_{1,\varepsilon_p}(x) + \varepsilon\sqrt{pn_p}P(A_{x,\varepsilon}) \\ = P(Y_p) &\geq \left.\frac{\delta_{1,\varepsilon_p}(x) - pn_p}{\sqrt{pn_p}} + \varepsilon\right) P(A_{x,\varepsilon}).\end{aligned}$$

which by Proposition 4.5 is bounded below, for sufficiently small p, by

$$P(Y_p \geq -x + \eta - \varepsilon)\, P(A_{x,\varepsilon}),$$

where η is an arbitrary positive number. hence by Lemmas 4.2 and 4.3,

$$\liminf_{p\to 0} P(N \leq z_p(x)) \geq \Phi(x - \eta + \varepsilon)$$

for all $\eta, \varepsilon > 0$, thereby completing the proof. □

Next we state and prove the upper bound counterpart to Lemma 4.3, namely,

Lemma 4.5. *For all x*

$$\limsup_{p\to 0} P(N \leq z_p(x)) \leq \Phi(x).$$

Proof. Clearly

$$P(N \leq z_p(x)) \leq P((N \leq z_p(x)) \cap A_{x,\varepsilon}) + P(A^c_{x,\varepsilon}). \tag{4.8}$$

Concerning the first term on the right hand side of (4.8), observe that

$$P((N \leq z_p(x)) \cap A_{x,\varepsilon}) \leq P((S_{n_p} \geq \gamma_\varepsilon(x) - \varepsilon(pn_p)^{1/2}) \cap A_{x,\varepsilon})$$

where γ_{ε_p} is as defined earlier. This time, the simpler function than γ_{ε_p} that we need must be smaller than γ_{ε_p}. To this end, define L_3 , L_4, and $\delta_{2,\varepsilon_p}(.)$ as follows: Let L_3 be the line with slope $C'(z_p(x))$ which passes through the point $(\xi_p,\ C(\xi_p))$. Let L_4 be the line with slope p which intersects L_3 at the point $z_p(x)$, and let $\delta_{2,\varepsilon_p}(.)$ be the function which maps x into the height of L_4 at $(1-\varepsilon_p)\xi_p$. Then $\delta_{2,\varepsilon_p}(z_p(x)) < \gamma_{\varepsilon_p}(z_p(x))$ and

$$\begin{aligned} P((N &\leq z_p(x)) \cap A_{x,\varepsilon}) \leq P((S_{n_p} \geq \delta_{2,\varepsilon_p}(x) - \varepsilon(pn_p)^{1/2}) \cap A_{x,\varepsilon}) \quad (4.9) \\ &= P\left(Y_p \geq \frac{\delta_{2,\varepsilon_p}(x) - pn_p}{(pn_p)^{1/2}} - \varepsilon\right) P(A_{x,\varepsilon}). \end{aligned}$$

The needed analogue to Proposition 4.4 is then

Proposition 4.6. *Given $\eta > 0$ there exists a p_0 such that*

$$\frac{\delta_{2,\varepsilon_p}(x) - pn_p}{\sqrt{pn_p}} > -x - \eta$$

for all $p > p_0$, and for all x.

Proof. The line L_3 satisfies the equation $y(w) - C(\xi_p) = C'(z_p(x))(w - \xi_p)$ and the equation of the line L_4 is $y(w) = pw - (pn_p - \delta_{2,\varepsilon_p}(x))$. At the point $z_p(x)$ the two lines intersect so

$$(C(\xi_p) + C'(z_p(x))(z_p(x) - \xi_p) = pz_p(x) - pn_p + \delta_{2,\varepsilon_p}(x))$$

or

$$\begin{aligned} \delta_{2,\varepsilon_p}(x) - pn_p &= p\xi_p - pz_p(x) + C'(z_p(x))(z_p(x) - \xi_p) \qquad (4.10) \\ &= -pxs_p(1 - C'(z_p(x))/p). \end{aligned}$$

However,

$$\begin{aligned} C'(z_p(x))/p &= A\rho(z_p(x))^{-(1-\rho)}/p = A\rho(\xi_p + xs_p)^{-(1-\rho)}/p \\ &= \rho(1 + xs_p/\xi_p)^{-(1-\rho)} = \rho(1 + Kp^{\frac{\rho}{2(1-\rho)}})^{-(1-\rho)} \end{aligned}$$

where K is a positive constant by (4.2). Thus (4.10) becomes

$$\delta_{2,\varepsilon_p}(z_p(x)) - pn_p = -pxs_p\{1 - \rho + O(p^{\frac{\rho}{2(1-\rho)}})\}.$$

Since $ps_p/\sqrt{pn_p} = 1/(1-\rho)\sqrt{1-\varepsilon_p}$, the proof is complete. □

The proof of Lemma 4.5 now follows straightforwardly from (4.8) and (4.9) upon application there of Proposition 4.6, Lemma 4.2, and the limit result of (4.2). □

Since Lemmas 4.3 and 4.5 together comprise Theorem 4.1, its proof is also complete.

Once we have the limiting distribution of N, a corollary gives us the limiting distribution of the estimator, $\overline{X}_n^a$.

Corollary 4.7. $p^{b/2}(\overline{X}_N^a - p^a) \xrightarrow{L} N(0, c)$ *as* $p \to 0$.

Proof. Since by construction, $C(N) \leq S_N < C(N) + 1$, and by (3.6), $C(N) = AN^\rho$, we obtain

$$AN^{\rho-1} \leq \overline{X}_N < AN^{\rho-1} + N^{-1}. \tag{4.11}$$

Write $Z_p = (N - \xi_p)/s_p$ as the random variable in Theorem 4.1 that converges to a standard normal. Direct substitution gives

$$\begin{aligned}
& p^{b/2}\{(C(N)/N)^a - p^a\} \\
= \;& p^{\frac{-\rho}{2(1-\rho)}}\{A^a(\xi_p + s_p Z_p)^{-a(1-\rho)}p^{-a} - 1\} \\
= \;& p^{\frac{-\rho}{2(1-\rho)}}\{(1 + (s_p/\xi_p)Z_p)^{-a(1-\rho)} - 1\}. \\
= \;& p^{\frac{-\rho}{2(1-\rho)}}\{1 - a(1-\rho)(s_p/\xi_p)Z_p + O(((s_p/\xi_p)Z_p)^2) - 1\} \\
= \;& c^{1/2}Z_p + O(p^{\frac{\rho}{2(1-\rho)}}Z_p^2)
\end{aligned}$$

where (4.2) has been used. Hence

$$p^{b/2}\{(C(N)/N)^a - p^a\} \xrightarrow{L} N(0, c) \text{ as } p \to 0.$$

Since $1/N$ is of smaller order than $C(N)/N = AN^{\rho-1}$, it is straightforward to show that $p^{b/2}\{(C(N)/N - 1/N)^a - p^a\}$ converges in law as well to the same $N(0, C)$ r.v., and thus completes the proof. □

5. Uniform Integrability and Convergence of the Risk.

In this section we obtain the key result about the uniform integrability of the loss function and the consequent convergence of the *risk*. In the process, we obtain a sharp bound on the tail probabilities for the loss function. To be able to establish these results, it is necessary to impose a minimum stopping time in order to offset the effect of an initial string of 'successes'. To this end introduce the lower limit

$$m_r = \inf\{k > 0 : C(k) > -\frac{r}{2}\min\{b, 2a + b\}\} \tag{5.1}$$

and in analogy to (3.4) and (3.5), define

$$N_r = \inf\{n \geq m_r : n \geq \xi_{\overline{X}_n}\} = \inf\{n \geq m_r : S_n \geq C(n)\}. \tag{5.2}$$

Let $Z_{p,r}$ denote the square root of the loss function, namely,

$$Z_{p,r} = p^{b/2}|\overline{X}^a_{N_r} - p^a|. \tag{5.3}$$

The key bound needed for this section is given as Theorem 5.1. For this, set

$$\tau_r = -\frac{r}{2}\min\{b, 2a+b\}, \qquad \gamma = \begin{cases} 1 & \text{if } a > 0 \\ \min(1, -1/2a(1-\rho)) & \text{if } a < 0. \end{cases} \tag{5.4}$$

Theorem 5.1. *For $\{Z_{p,r} : p \in (0,1)\}$ as defined in (5.3), there exist positive constants K, K_1, K_2, K_3, $\tau > \tau_r$ and $0 < \epsilon < 1$ such that for all $u \geq 0$;*

i. $P(Z_{p,r} > u, N_r < \xi_p) \leq K_1 p^\tau + K_2 e^{-K_3 u}$ *for p sufficiently small, and*

ii. $P(Z_{p,r} > u, N_r \geq \xi_p) \leq e^{-Ku^{2\gamma}}$ *for all $p \in (0,1)$,*

where τ_r and γ are defined in (5.4).

We delay discussion of the proof of these bounds until later. We state next the uniform integrability result and give its proof using Theorem 5.1.

Theorem 5.2. *For any $p_0 \in (0,1)$ and $r > 0$, the family $\{Z^r_{p,r} : 0 < p \leq p_0\}$ is uniformly integrable.*

Proof. It suffices to show that $E(Z^s_{p,r}) < \infty$ for some $s > r$. To this end, write

$$\begin{aligned} E(Z^s_{p,r}) &= \int_0^\infty P(Z_{p,r} > u)du^s \qquad (5.5) \\ &= \int_0^\infty P(Z_{p,r} > u, N_r < \xi_p)du^s + \int_0^\infty P(Z_{p,r} > u, N_r \geq \xi_p)du^s. \end{aligned}$$

From Theorem 5.1, it follows that only the first integral needs attention, and that is because of the constant term, $K_1 p^\tau$, that is part of its relevant bound. However, the range of integration is bounded. To see this, note first that by (3.2), $N_r < \xi_p$ is equivalent to $AN_r^{\rho-1} > p$. However, by (3.6), $C(N_r)/N_r = AN_r^{\rho-1}$. Also, by definition of the stopping time N_r as a first passage time,

$$\frac{C(N_r)}{N_r} \leq \overline{X}_{N_r} < \frac{C(N_r)}{N_r} + \frac{1}{N_r}; \tag{5.6}$$

cf. (4.11). Thus on the event $(N_r < \xi_p)$, it follows that $\overline{X}_{N_r} > p$, and so for $a > 0$ the event in the first integral of (5.5) is contained in the event $(p^{b/2}(\overline{X}^a_N - p^a) > u)$, an event that is empty unless $u \leq p^{b/2}$. Hence by Theorem 5.1, for sufficiently small p the first integral is bounded above by

$$\int_0^{p^{b/2}} \{K_1 p^\tau + K_2 e^{-K_3 u}\} du^s$$

which is uniformly bounded provided only that $\tau + sb/2 \geq 0$. However, since $\tau > \tau_r \equiv -rb/2$ when $a > 0$, any $s \in (r, -2\tau/b]$ would suffice.

Similarly, for $a < 0$, $N_r < \xi_p$ implies that $\overline{X}^a_N < p^a$, so that the event in the first integral of (5.5) is now contained in the event $(p^{b/2}(p^a - \overline{X}^a_N) > u)$, which is empty unless $u \leq p^{a+b/2}$. Thus, in this case the integral will be bounded if $\tau + s(a + b/2) \geq 0$. But now $\tau > \tau_r \equiv -r(a + b/2)$, allowing for any $s \in (r, -\tau/(a_b/2)]$. Thus for either case, use any $s \in (r, r\tau/\tau_r]$. □

The main consistency result of this paper, described in (1.12), is now an immediate consequence of the uniform integrability established in Theorem 5.2.

Corollary 5.3. (Consistency) *For any $r \geq 2$, the sequential estimation procedures based on the stopping time N_r satisfy*

$$\lim_{p \to 0} \text{Risk} \equiv \lim_{p \to 0} E(Z^2_{p,r}) = c.$$

Remarks on the Proof of Theorem 5.1. The derivation of the bounds in Theorem 5.1 requires lengthy computations that rely heavily upon the power form of the functions g and h in our loss function; see (3.1). Moreover, the steps in the proof follow closely those detailed in Hubert and Pyke (1997) for the continuous Brownian model. In fact, it was the methodology used to solve the Brownian motion problem that pointed the way to the proofs in the more difficult discrete problem of this paper. The self-similarity of Brownian motion enables one to transform the original problem that involves a single stopping curve applied to a family of processes (indexed by p) into a problem involving a family of stopping curves applied to a single process. In the transformed problem, the appropriate linear approximations to the curves were more readily determined. As for the continuous case, the proof here would require a partitioning of the range of N_r into three intervals, and then approximation of the curve C by a straight line over each of these intervals.

For the discrete problem of this paper, an added difficulty is the jump over the boundary at N_r, as seen in (5.6). To handle this, the approach is to establish first the probability bounds of Theorem 5.1 for the approximate square-root loss given by

$$Z^*_{p,r} \equiv p^{b/2} |(C(N_r)/N_r)^a - p^a|, \tag{5.7}$$

and then deduce therefrom the bounds for $Z_{p,r}$. The particular details, though not the ideas, are different for the two cases of $a > 0$ and $a < 0$.

Suppose $a > 0$. Then, to split the range of the variable N_r into three pieces and approximate the curve C with a straight line on each piece, we begin as follows. Observe

$$\begin{aligned}[][Z^*_{p,r} > u, N_r < \xi_p] &= [p^{b/2}|(C(N_)/N_r)^a - p^a| > u, N_r < \xi_p] \\ &= [p^{b/2}\{(C(N_r)/N_r)^a - p^a\} > u] \\ &= (AN_r^{\rho-1})^a > p^a(1 + up^{\rho/2(1-\rho)}] \qquad (5.8) \\ &= [N_r < (A/p)^{1/(1-\rho)}(1 + up^{\rho/2(1-\rho)})^{-1/a(1-\rho)}] \\ &= [N_r < \xi_p t_{p-}(u)]\end{aligned}$$

where $t_{p-}(u) = (1 + up^{\rho/2(1-\rho)})^{-1/a(1-\rho)}$ for $a > 0$. We therefore need to bound $P(N_r < \xi_p t_{p-}(u)) \equiv P_p(u)$, say, for each u. To this end, choose $\varepsilon \in (0, 1)$ and set

$$v_p = Kp^{(1-\varepsilon)/2(1-\rho)}$$

where K is a positive constant. Notice that the maximum value of $C(t) - tp$ occurs at the point $t_{\max} = \rho^{1/(1-\rho)}\xi_p$ so that $\xi_p v_p$ is to the left of $t_{\max}$ for p sufficiently small. Let w_p be the unique value greater than $t_{\max}/\xi_p$ which satisfies

$$C(\xi_p v_p) - \xi_p v_p p = C(\xi_p w_p) - \xi_p w_p p.$$

Define u_1 and u_2 by

$$t_{p-}(u_1) = v_p, \qquad \text{and} \qquad t_{p-}(u_2) = w_p.$$

The range of u is split first into the three intervals I, II, and III defined by

$$\text{I. } u_1 < u, \qquad \text{II. } u_2 < u \le u_1, \qquad \text{III. } 0 \le u \le u_2.$$

Although these intervals are described for u, their resulting partitions of the n-axis through $\xi_p t_{p-}(u)$ are the more natural, and should be used by the reader in drawing figures to assist with the reading. Consider interval I. Observe first that since N_r is bounded below by m_r, we have $P(N_r < \xi_p t_{p-}(u)) = 0$ whenever u is such that $t_{p-}(u)\xi_p < m_r$. Consider therefore only those u for which $t_{p-}(u)\xi_p \ge m_r$. Unlike for the continuous problem we are unable to get a single bound that applies over the entire interval I, but must further split I into subintervals, I_a and I_b, say, as follows. Define u_3 by $\xi_p t_{p-}(u_3) = m'$ where m' is a constant greater than m_r to be fixed later. Then define $\text{I}_a = (u_3, \infty)$ and $\text{I}_b = (u_1, u_3]$.

For u's in interval I_a, $N_r < \xi_p t_{p-}(u)$ implies that $S_n \geq C(n)$ for some $m_r \leq n < m'$. The probability that S_n crosses $C(n)$ in I_a is bounded by the probability that S_n crosses the horizontal line, L_0, of height $C(m_r)$ before m'; that is

$$\begin{aligned} P(N_r < \xi_p t_{p-}(u)) &\leq P(N_r < \xi_p t_{p-}(u_3)) = P(N_r < m') \\ &\leq P(S_{m'} \geq C(m_r)) < Kp^{\lceil C(m_r) \rceil}. \end{aligned} \tag{5.9}$$

We note that (5.9) is the reason we defined m_r as in (5.1). It should also be noted that it is easy to show that this is the *smallest* that m_r can be if Theorem 5.2 is to be true.

On interval I_b we bound the curve from below with the line L_1 that passes through the points $(m', C(m'))$ and $(\xi_p v_p, C(\xi_p v_p))$. We use a result due to Täcklind (1942) to obtain a suitable bound on the probability of crossing L_1. Denote the slope and intercept of L_1 by α_p and β_p, respectively. Let T_k denote the time of the k^{th} success, that is, $T_k = \inf\{j > 0 : S_j = k\}$. Our process S_n increases only at n's which coincide with a T_k, so the time of crossing of L_1 must be equal to T_k for some k. S_n will cross at time T_k if and only if $k \geq \beta_p + \alpha_p T_k$, or equivalently, if and only if $T_k - (k - \beta_p)/\alpha_p \leq 0$. The independent and equidistributed increments for Täcklind's theorem are $T_k - T_{k-1} - 1/\alpha_p$, and their generating function is

$$\phi(s) = E(e^{s(T_k - T_{k-1} - 1/\alpha_p)} = pe^{-s(1/\alpha_p - 1)}(1 - qe^s)^{-1}.$$

According to Täcklind's result we have

$$P(\text{absorption}) \leq e^{-R\beta_p/\alpha_p}$$

for any $R > 0$ such that $\phi(-R) \leq 1$. It may be shown that $R \equiv R_p \equiv -k_0 \alpha_p \ln(p)/\beta_p$ may be used where k_0 is a positive constant. In fact, the calculation of $\phi(-R_p)$ shows that it converges to 0 as $p \to 0$. Therefore,

$$P(N_r < \xi_p t_{p-}(u)) \leq P(\text{absorption}) \leq e^{-R_p \beta_p/\alpha_p} = e^{k_0 \ln(p)} = p^{k_0}$$

for $u_1 < u \leq u_3$ and for p sufficiently small. Choose k_0, then m', and then ε such that

$$k_0 > -rb/2, C(m') > 2k_0, \text{ and } 1/2 - \varepsilon/2 - k_0/C(m') \geq 0.$$

We can then combine (5.1) with this to get

$$P(N_r < \varepsilon_p t_{p-}(u)) \leq Kp^{\tau_r} \tag{5.10}$$

for any $u > u_1$, for p sufficiently small, where $\tau_r = \min(k_0, \lceil C(m_r) \rceil) > -rb/2$.

Next, consider interval II. For this interval we use the same bound on the interval I section and a line, L_2, which has slope p and passes through the point $(\xi_p v_p, C(\xi_p v_p))$ on the interval II section. The probability of crossing the boundary formed by L_0, L_1, and L_2 between m_r and $\xi_p t_{p-}(u)$ is less than or equal to the probability bound in (5.10) plus the probability of crossing L_2 between 0 and ξ_p. By means of the Skorokhod and Bernstein Inequalities, we obtain

$$P(\max_{1\leq k\leq \xi_p} |S_k - kp| \leq \delta) \leq e^{-\delta^2/2(\xi_p\sigma^2+\delta/3)} \tag{5.11}$$

where $\sigma^2 = pq$ and

$$\begin{aligned}\delta = C(\xi_p v_p) - \xi_p v_p p &= K' p^{-\rho(1+\varepsilon)/2(1-\rho)} - K'' p^{-1(1+\varepsilon)/2(1-\rho)+1} \\ &= p^{-\rho(1+\varepsilon)/2(1-\rho)}(K' - K'' p^{(1-\varepsilon)/2}) = O(p^{-\rho(1+\varepsilon)/2(1-\rho)}).\end{aligned}$$

Hence, $\xi_p\sigma^2 + \delta/3 = O(p^{-\rho/(1-\rho)})$, leading to

$$\begin{aligned}P(N_r < \xi_p t_{p-}(u)) &\leq e^{-Kp^{\frac{-\rho(1+\varepsilon)}{(1-\rho)}+\frac{\rho}{1-\rho}}} + Kp^{\tau_r} \\ &= e^{-Kp^{\frac{-\varepsilon\rho}{1-\rho}}} + Kp^{\tau_r},\end{aligned} \tag{5.12}$$

for suitable K whenever $u_2 < u \leq u_1$.

For interval III the method is similar, but the L_2 line is replaced by the line L_3 which is below L_2, has slope p, and passes through $(\xi_p t_{p-}(u), C(\xi_p t_{p-}(u))$. Whereas L_2 depends only on p, L_3 depends on u as well as on p. Once again we use the bound in (5.10) for the interval I section of the bound. We again use the Skorokhod and Bernstein Inequalities to obtain (5.11) but this time with

$$\begin{aligned}\delta &= C(\xi_p t_{p-}(u)) - \xi_p t_{p-}(u)p \\ &= A^{1/(1-\rho)}\{p^{-\rho/(1-\rho)} t^{\rho}_{p-}(u) - p^{-1/(1-\rho)} t_{p-}(u)p\} \\ &= Kp^{-\rho/(1-\rho)} t^{\rho}_{p-}\{1 - (1 + up^{\rho/2(1-\rho)})^{-1/a}\} \\ &= Kp^{-\rho/2(1-\rho)} t^{\rho}_{p-}(u)u\{1 + O(up^{\rho/2(1-\rho)})\}.\end{aligned}$$

It may be shown as in Hubert and Pyke (1997) that $up^{\rho/(1-\rho)} \to 0$ and $t_{p-}(u) \to 1$ over $0 \leq u \leq u_2$, from which it follows that $\delta = O(p^{\rho/2(1-\rho)}u)$. It follows then that the probability of crossing L_3 satisfies

$$\begin{aligned}P(\text{cross } L_3) &\leq \exp\{\tfrac{-O(u^2 p^{-\rho/(1-\rho)})}{O(p^{-\rho/(1-\rho)}) + O(up^{-\rho/2(1-\rho)})}\} \\ &\leq e^{-K\min(O(u^2), O(up^{-\rho/2(1-\rho)}))} \leq e^{-Ku}\end{aligned} \tag{5.13}$$

for p sufficiently small. In view of (5.10), this leads to

$$P(N < \xi_p t_{p-}(u)) \leq K_1 p^{\tau_r} + e^{-K_2 u} \tag{5.14}$$

for $0 \le u \le u_2$, and p sufficiently small. To simplify, we combine the three bounds in (5.10), (5.12), and (5.14) to get

$$P(N < \xi_p t_{p-}(u)) \le K_1 p^{\tau_r} + e^{-K_2 p^{-\varepsilon\rho/(1-\rho)}} + e^{-K_3 u} = K_4 p^{\tau_r} + e^{-K_5 u} \tag{5.15}$$

for $u \ge 0$ and p sufficiently small.

Consider now the right hand tail of N_r in the case $a > 0$. The set

$$\begin{aligned} [Z^*_{p,r} > u, N_r \ge \xi_p] &= [p^{b/2}|C(N_r)/N_r)^a - p^a| > u, N - r \ge \xi_p] \\ &= [p^{b/2}(p^a - (C(N_r)/N_r)^a) > u] \\ &= [N_r > (A/p)^{1/(1-\rho)}(1 - up^{\rho/2(1-\rho)}))^{-1/a(1-\rho)} \\ &= [N_r > \xi_p t_{p+}(u)] \end{aligned} \tag{5.16}$$

where $t_{p+}(u) = (1 - up^{\rho/2(1-\rho)})^{-1/a(1-\rho)}$. It is important to note from the second line in the above expression that always $u < p^{a+b/2} \equiv p^{-\delta/2(1-\delta)}$. Now $N_r > \xi_p t_{p+}(u)$ implies that the partial sums S_n have remained below the stopping curve for $n = m_r, \dots, \xi_p t_{p+}(u)$ which implies in particular that $S_{\xi_p t_{p+}(u)} < C(\xi_p t_{p+}(u))$. (Note that we take a u which gives us an integer value of $\xi_p t_{p+}(u)$ here.) Thus,

$$\begin{aligned} P(N_r > \xi_p t_p(u)) &\le P(S_{\xi_p t_{p+}(u)} C(\xi_p t_{p+}(u))) \\ &= P\left(\frac{S_n - np}{n} < \frac{C(n) - np}{n}\right) \end{aligned}$$

where $n = \xi_p t_{p+}(u)$. Bennett's Inequality with $n = \xi_p t_{p+}(u)$, $t = (C(n) - np)/n$, $\sigma^2 = pq$ and $b = -p$ states that

$$P\left(\frac{S_n - np}{n} < t\right) \le \exp\{-\frac{nt}{b}\{(1 + \frac{\sigma^2}{bt}) \quad \ln\left(1 + \frac{bt}{\sigma^2}\right) - 1\}\}$$

for $-p < t < 0$. An expansion of the logarithm inside the braces, together with the fact that bt/σ^2 is positive, leads to

$$P(N_r > \xi_p t_{p+}(u)) \le e^{-K\frac{nt}{b}\frac{bt}{\sigma^2}} = e^{-Knt^2/\sigma^2}. \tag{5.17}$$

Upon substituting $x = up^{\rho/2(1-\rho)}$ we get

$$\begin{aligned} t_{p+}(u) &= (1-x)^{-1/a(1-\rho)}, \qquad n = \xi_p(1-x)^{-1/a(1-\rho)}, \text{ and} \\ t &= \frac{C(n)}{n} - p = An^{-(1-\rho)} - p \\ &= A\xi_p^{-(1-\rho)}(1-x)^{1/a} - p = p((1-x)^{1/a} - 1). \end{aligned}$$

From (5.17) we then obtain, making use of the fact that $p^{-\rho(1-\rho)} = u^2/x^2$,

$$\begin{aligned} P(N_r > \xi_p t_{p+}(u)) &\leq \exp\{-K\xi_p(1-x)^{-1/a(1-\rho)}p^2\{(1-x)^{1/a}-1)^2\}/pq\} \\ &= \exp\{-Kp^{-\rho/(1-\rho)}(1-x)^{-1/a(1-\rho)}\{1-(1-x)^{1/a}\}^2\} \\ &\leq e^{-Ku^2 J(x)} \end{aligned} \tag{5.18}$$

where

$$J(x) = x^{-2}(1-x)^{-1/a(1-\rho)}\{1-(1-x)^{1/a}\}^2.$$

Because of the form of t_{p+} and the remark following (5.16), x is always between 0 and 1 for all values of p and u to be considered. It follows then that J must have a strictly positive minimum. From that and (5.18) we conclude that

$$p(N_r > \xi_p t_{p+}(u)) \leq e^{-Ku^2} \tag{5.19}$$

for all $p \in (0,1)$ and $u > 0$.

For the case of $a < 0$, the steps in the proof are similar to the above. The definition of $t_{p-}(u)$ changes to $t_{p-}(u) = (1-up^{\rho/2(1-\rho)})^{-1/a(1-\rho)}$ so that $[Z^*_{p,r} > u, N_r < \xi_p] = [N_r < \xi_p t_{p-}(u)]$. The range of u, and hence of n, is split into three intervals as before. Intervals I and II are similar since they depend only on the curve C and on the values m_r, m', v_p, w_p, and ξ_p. Interval I_a is exactly the same so we get the same result as in (5.7), namely,

$$P(N_r < \xi_p t_{p-}(u)) < Kp^{\lceil C(m_r)\rceil}$$

for u in I_a. For I_b, proceed as before but choose k_0, ε, and m' such that

$$k_0 > -r(2a+b)/2, \quad C(m') > 2k_0, \text{ and } 1/2 - \varepsilon/2 - k_0/C(m') > 0,$$

for which

$$P(N_r < \xi_p t_{p-}(u)) \leq p^{k_0} \tag{5.20}$$

for $u_1 < u \leq u_3$ and p sufficiently small. With these changes the conclusions for u in intervals I and II are

$$\begin{aligned} P(N_r < \xi_p t_{p-}(u)) &\leq Kp^{\tau_r}; u_1 < u \\ P(N_r < \xi_p t_{p-}(u)) &\leq e^{-K_1 p^{-\varepsilon\rho/(1-p)}} + K_2 p^{\tau_r}; \qquad u_2 < u \leq u_1 \end{aligned}$$

where $\tau_r = \min(k_0, \lceil C(m_r)\rceil) > -r(2a+b)/2$. Interval III is nearly the same as it was for $a > 0$; the only difference is in the definition of t_{p-}. Again, in Bernstein's Inequality

$$\begin{aligned} \delta &= C(\xi_p t_{p-}(u) - \xi_p t_{p-}(u)p) \\ &= A^{1/(1-\rho)}p^{-\rho/(1-\rho)}t^{\rho}_{p-}(u)(1-t^{1-\rho}_{p-}(u)) \leq Kp^{-\rho/2(1-\rho)}u \end{aligned}$$

for p sufficiently small, so that (5.13), and hence (5.14) and (5.15), hold also when $a < 0$.

For the right hand tail of N_r when $a < 0$, t_{p+} is defined by $t_{p+}(u) = (1 + up^{\rho/2(1-\rho)})^{-1/a(1-\rho)}$. Proceed as before to (5.17), the bound obtained from Bennett's Inequality. The substitution $x = up^{\rho/2(1-\rho)}$ now leads to quantities in which $1 + x$ replaces the former $1 - x$, namely,

$$\begin{aligned} t_{p+}(u) &= (1+x)^{-1/a(1-\rho)}, \quad n = \xi_p(1+x)^{-1/z(1-\rho)}, \text{ and} \\ t &= \frac{C(n)}{n} - p = An^{-(1-\rho)} - p = p\{(1+x)^{1/a} - 1\}. \end{aligned}$$

Since $0 < x/u < 1$ by definition of x, we have $p^{-\rho/(1-\rho)} \equiv u^2x^{-2} \geq u^{2\varepsilon}x^{-2\varepsilon}$ for any $0 < \varepsilon \leq 1$. thus, the bound's exponent satisfies

$$\begin{aligned} \frac{nt^2}{\sigma^2} &\equiv \xi_p(1+x)^{-1/a(1-\rho)}p^2\{(1+x)^{1/a} - 1\}^2(pq)^{-1} \\ &\geq A^{1/(1-\rho)}p^{-\rho/(1-\rho)}(1+x)^{-1/a(1-\rho)}\{(1+x)^{1/a} - 1\}^2 \\ &\geq Ku^{2\varepsilon}x^{-2\varepsilon}(1+x)^{-1/a(1-\rho)}\{(1+x)^{1/a} - 1\}^2, \end{aligned}$$

for all $0 < \varepsilon \leq 1$. Here, x takes on all positive values in contrast to the previous case of $a > 0$. We check first that when $\varepsilon = 1$

$$\lim_{x \to 0} x^{-2}(1+x)^{-1/a(1-\rho)}\{(1+x)^{1/a} - 1\}^2$$

is positive. Moreover, if $-1/a(1-\rho) \geq 2$, then the limit as $x \to \infty$ is also either positive or $+\infty$. However, to cover the other case, choose $\varepsilon = \gamma \equiv \min\{1, -1/2a(1-\rho)\}$ for which the limit inferior of $nt^2\sigma^{-2}$ as $x \to \infty$ exceeds $Ku^{2\varepsilon}$. Since the two limits are bounded away from zero, the function is continuous, and the function is always positive; it must have a strictly positive minimum, enabling us to conclude

$$P(N_r > \xi_p t_{p+}(u)) \leq e^{-Ku^{2\gamma}} \tag{5.21}$$

for $u > 0$ and appropriate constants K and $0 < \varepsilon < 1$. □

This completes the derivation of the bounds of Theorem 5.1 for the $Z^*_{p,r}$ of (5.7). It would remain to show that hey apply as well to the desired $Z_{p,r}$ of (5.3). The only difference between the two is that $Z^*_{p,r}$ uses $C(N_r)/N_r$ to estimate p, whereas $Z_{p,r}$ uses $\overline{X}_{N_r}$. Recall from (5.6) that the difference is small. However, it requires considerable detail to handle it adequately, detail that we will not give; the reader will find one approach in Hubert (1986). the setting, however, can be clarified by introducing the simplifying notation $X_p = p^{-1}\overline{X}_{N_r}$ and $Y_p = p^{-1}AN_r^{\rho-1}$ in terms of which (5.3), (5.6) and (5.7) become

$$\begin{aligned} Y_p &\leq X_p < Y_p + (pY_p/A)^{1/(1-\rho)} \\ Z^*_{p,r} &= p^{-\rho/(1-\rho)}|Y_p^a - 1|, \quad Z_{p,r} = p^{-\rho/(1-\rho)}|X_p^a - 1|. \end{aligned} \tag{5.22}$$

In this notation, the event $(N_r < \xi_p) = (Y_p > 1) \subset (X_p > 1)$, while the event $(N_r > \xi_p) = (Y_p < 1) \subset (X_p < 1 + (p/A)^{1/(1-\rho)})$. It is again necessary to separate the cases of $a > 0$ and $a < 0$, but in both cases one handles the difference $Z^*_{p,r} - Z_{p,r}$ through the inequalities of (5.22) and applying the bounds already obtained for Y_p to control the difference $(X_p - Y_p) \leq (pY_p/A)^{1/(1-\rho)}$. □

6. Remarks and Open Problems. In this paper, we have presented a new approach to the asymptotic analysis of the sequential estimation problem for functions of the Binomial parameter p. Instead of letting the goal of the problem (such as the accuracy or the cost per observation) change in order to drive the asymptotics, this approach leaves the goal unchanged, focusing instead upon asymptotics in which the parameter approaches zero. Within this different framework we have given conditions under which the procedures are efficient with respect to having the random sample sizes close to the best fixed sample sizes (Section 3), shown that the estimator and sample size are asymptotically normal (Section 4) and obtained essentially exponential tail bounds for the sample size that yield uniform integrability of the loss functions and the consequent consistency of the estimator and convergence of the risk (Section 5). The proofs of these results depend significantly upon the particular power form of the functions g and h in the loss function of (3.1).

It would be of interest to study this problem for more general g and h. To see how such a generalization might proceed, consider (1.1)–(1.4) and the start of the proof of Theorem 3.1. If we follow the same procedure and put off the specialization until later we come up with the stopping time $N = \inf\{n > 0 : S_n \geq n\xi^{-1}(n)\}$ where $\xi_p = h(p)g'(p)^2p/c$ and ξ^{-1} is the inverse of ξ_p, namely, $\xi^{-1}(\xi_p) = p$.

In order for the definition of N to make sense we take ξ_p to be a monotone, decreasing function of p, so that the ideal expected sample size, ξ_p, increases as $p \to 0$, which certainly is reasonable in the context of our problem.

Presumably, if $n\xi^{-1}(n)$ is increasing, concave, and sufficiently smooth, (as it is in this paper when $2a + b < 0$) then similar theorems to those that we have proved may still hold. However, the methods of proof used here are not easily adaptable to this more general case; we make essential use of the fact that we could explicitly invert ξ_p and solve certain equalities and inequalities that arise. The outline of the proofs may well carry over to the more general situation, but the details appear to become substantially more difficult. One particular special case that is easy is the case $h(p) = 1$ and $g(p) = \log(p)$. In this case, $n\xi^{-1}(n)$ is constant, so the problem is very similar to that of the boundary case, $2a + b = 0$, when g and h are powers of p. As in that case, N_p/ξ_p does not converge to 1 (see the discussion prior to (1.10)).

Several aspects of our approach to the problem have been ad hoc in nature, though they are consistent with work in the literature. For example, we have arbitrarily chosen a loss function which is a function of p multiplied by squared error loss. The loss could clearly take a more general form than this. The fact that the loss was squared did not play an important role and the proofs would probably go through with little modification for other powers. The multiplicative nature of the loss was useful as a simplifying feature, thought it is not clear that one would choose something more general.

Another ad hoc feature of the formulation is that we use $g(\overline{x})$ to estimate $g(p)$. This is done to simplify the problem and to make our results comparable to other results in the literature. A more general approach would be one in which the problem is stated in terms of a general estimator, t_n, and the actual form of t_n becomes one of the outputs of the analysis.

An Empirical Model. An important feature of the continuous problem studied in Hubert and Pyke (1997) was that it was possible to describe the entire family of stopping problems, indexed by p, in terms of a single random process. This application of the self-similarity of Brownian Motion resulted in a simpler analysis, which in turn suggested the approach used here in Section 5 for the discrete problem even though the Bernoulli random walks do not have this self-similarity property. There is, however, a way to embed all of the stopping problems for the Binomial case into a single process based upon a sequence of uniform rv's. This embedding is described as follows.

Let $U_1, U_2, \ldots$ be iid uniform $(0,1)$ rv's. View each of the U_i's plotted at (U_i, i) so that one is on each of the dashed lines in Figure 6.1. Let $S_{n,p}$ be the number of U_i's out of the first n that are less than p, that is,

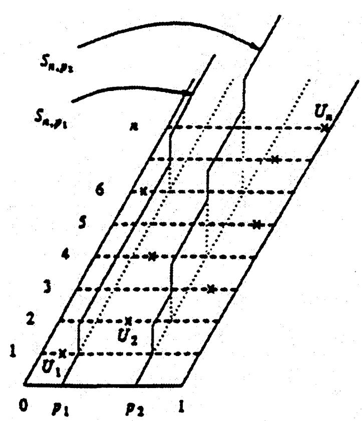

Figure 6.1

$$S_{n,p} = \sum_{i=1}^{n} 1_{[U_i \le p]}. \tag{6.1}$$

Let C be the same function as defined in (3.6). The stopping time $N^{(p)}$ is defined as

$$N^{(p)} = \inf\{n > 0 : S_{n,p} \ge C(n)\}. \tag{6.2}$$

This is the same as the stopping time N defined in (3.5) where there the parameter p was suppressed from the notation. With this formulation we have defined the entire family of stopping problems on the single random process $\{U_n; n \ge 1\}$. We think of the curve $C(n)$ as a roof over the process (see Figure 6.2). Notice that C does not depend on p. The value of $N^{(p)}$ is the value of n at the point where for the first time $S_{n,p} \ge C(n)$. With the coupling implicit in this formulation, $N_{p_1} \ge N_{p_2}$ for $p_1 < p_2$. In the previous sections, we study $N^{(p)}$ as $p \to 0$, utilizing, of course, only the marginal distributions of the $N^{(p)}$. However, the formulation given here raises interesting questions about a stopping 'curve' determined by the first crossings of a surface by a random surface. Here, for example, the set of $N^{(p)}$'s form a curve in the (n, p) plane; see Figure 6.3. It would be of interest in particular to determine in this model, the a.s. rate at which this curve approaches $+\infty$ as $p \to 0$.

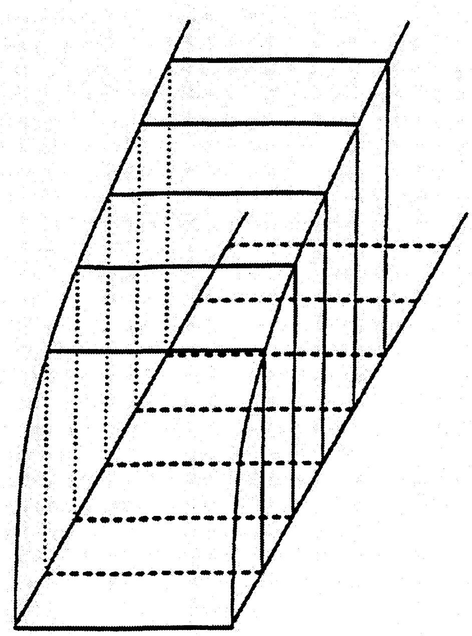

Figure 6.2

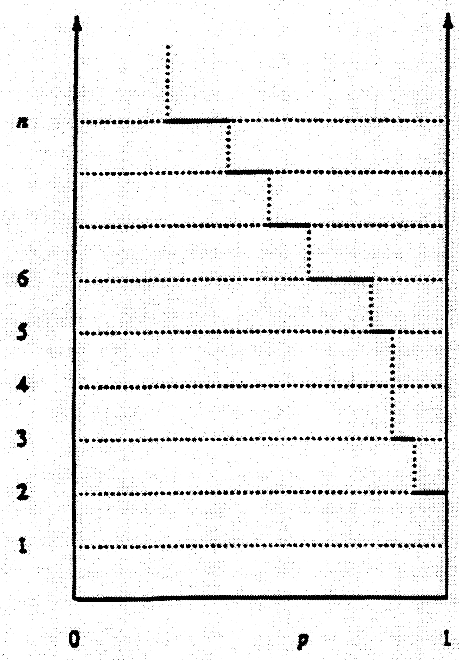

Figure 6.3

References

Alvo, M. (1977). Bayesian sequential estimation. *Ann. Statist.* **5**, 955–968.

Anscombe, F. J. (1953). Sequential estimation. *J. Roy. Stat. Soc.* Ser B **15**, 1–21.

Cabilio, P. (1977). Sequential estimation in Bernoulli trials. *Ann. Statist.* **5**, 342–356.

Cabilio, P. and Robbins, H. (1975). Sequential estimation of p with squared relative error loss. *Proc. Nat. Acad. Sci. USA* **72**, 191–193.

Chow, Y. S. and Robbins, H. (1965). On the asymptotic theory of fixed width sequential confidence intervals for the mean. *Ann. Math. Statist.* **36**, 457–462.

Hájek, J. and Rényi, A. (1955). Generalization of an inequality of Kolmogorov, *Acta Math. Acad. Sci. Hungar.* **6**, 281–283.

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

Haldane, J. B. S. (1956). the estimation and significance of the logarithm of a ratio of frequencies. *Ann. Human Genetics* **20**, 309–311.

Hubert, S. L. (1986). Sequential estimation of functions of p for Bernoulli trials. Ph.D. dissertation, Univ. of California, Berkeley.

Hubert, S. L. and Pyke, R. (1997). A particular application of Brownian motion to sequential analysis. *Statist. Sinica* **7**, 109–126.

Robbins, H. (1959). Sequential estimation of the mean of a normal population. In *Probability and Statistics (The Harold Cramér Volume)*, Almqvist and Wiksell, Uppsala, 235–245.

Robbins, H. and Siegmund, D. (1970). Boundary crossing probabilities for the Wiener process and sample sums, *Ann. Math. Statist.* **41**, 1410–1429.

Robbins, H. and Siegmund, D. (1974). Sequential estimation of p in Bernoulli trials. *Studies in Probability and Statistics*, (papers in honour of Edwin J. G. Pitman), edited by E. J. Williams, North-Holland, Amsterdam, 103–107.

Siegmund, D. (1982). A sequential confidence interval for the odds ratio. *Probab. Math. Statist.* **2**, 149–156.

Simons, G. (1968). On the cost of not knowing the variance when making a fixed-width confidence interval for the mean. *Ann. Math. Statist.* **39**, 1946–1952.

Starr, N. (1966a). The performance of a sequential procedure for the fixed width interval estimation of the mean. *Ann. Math. Statist.* **37**, 37–50.

Starr, N. (1966b). On the asymptotic efficiency of a sequential procedure for estimating the mean. *Ann. Math. Statist.* **37**, 1173–1185.

Starr, N. and Woodroofe, M. (1968). Remarks on a stopping time. *Proc. Nat. Acad. Sci. USA* **61**, 1215–1218.

Starr, N. and Woodroofe, M. (1969). Remarks on sequential point estimation. *Proc. Nat. Acad. Sci. USA* **62**, 285–288.

Starr, N. and Woodroofe, M. (1972). Further remarks on sequential point estimation: the exponential case. *Ann. Math. Statist.* **43**, 1147–1154.

Stein, C. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.* **16**, 243–258.

Täcklind, S. (1942). Sur le risque de ruine dans des jeux inéquitables. *Skand. Aktuarietidskrift* **25**, 1-42.

Vardi, Y. (1979a). Asymptotic optimality of certain sequential estimators. *Ann. Statist.* **7**, 1034–1039.

Vardi, Y. (1979b). Asymptotic optimal sequential estimation: the Poisson case. *Ann. Statist.* **7**, 1040–1051.

Wald, A. (1951). Asymptotic minimax solutions of sequential point estimation problems. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley and Los Angeles, 1–11.

Woodroofe, M. (1977). Second order approximations for sequential point and interval estimation. *Ann. Statist.* **5**, 984–995.

Woodroofe, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*, Society for Industrial and Applied Mathematics, Philadelphia.

NETWORKS AND DISTRIBUTED COMPUTING, BOX 354841, UNIVERSITY OF WASHINGTON, SEATTLE, WA 98195–4841
Electronic mail: HUBERT@CAC.WASHINGTON.EDU

DEPARTMENT OF MATHEMATICS, BOX 354350, UNIVERSITY OF WASHINGTON, SEATTLE, WA 98195–4350
Electronic mail: PYKE@MATH.WASHINGTON.EDU
(206) 543–1725; Fax (206) 543–0397