The Annals of Statistics
1983, Vol. 11, No. 1, 68–77

SEQUENTIAL SAMPLING BASED ON THE OBSERVED FISHER INFORMATION TO GUARANTEE THE ACCURACY OF THE MAXIMUM LIKELIHOOD ESTIMATOR

BY PATRICIA GRAMBSCH

Bell Laboratories

A sequential sampling scheme in which observations are taken until the observed Fisher information exceeds a target value is considered for a restricted class of parametric families. It is proved that the theorems for fixed sample size asymptotic likelihood inference conditional on ancillary statistics are directly applicable to the sequential sampling scheme for location families. The fixed sample-size approximate ancillary statistic for two-dimensional curved exponential families is shown to be ancillary in the sequential sampling scheme. Evidence from Monte Carlo simulations suggests the applicability of the fixed sample size conditional likelihood inference theorems to these families as well. The distribution of the sequential sample size is shown to be asymptotically normal.

1. Introduction. A recent series of papers (Efron and Hinkley, 1978; Hinkley, 1980; Cox, 1980) has used asymptotic theory for likelihood functions to show how approximate conditional inference can be made for single parameter models. Exactly or approximately ancillary statistics are used to partition the sample space into reference subsets for inferential probability calculations. An ancillary statistic is a function of the (minimal) sufficient statistic whose distribution does not depend on the value of the parameter. Thus far, the discussions have dealt with fixed sample size experiments.

Suppose that $X_1, \cdots, X_n$ are independent and identically distributed (iid), with probability density $f_\theta(\mathrm{x})$ where θ is a scalar. With $s_n = (x_1, \cdots, x_n)$, the log of the likelihood function is

$$\ell_\theta(s_n) = \sum \ln f_\theta(x_i).$$

The first two derivatives with respect to θ may be denoted by $\dot{\ell}_\theta(s_n)$ and $\ddot{\ell}_\theta(s_n)$ and abbreviated $\dot{\ell}_\theta$ and $\ddot{\ell}_\theta$. Under suitable regularity conditions, the maximum likelihood estimate (MLE) $\hat{\theta}_n$ is a solution of $\dot{\ell}_\theta = 0$. In addition to the *a priori* expected Fisher information

$$\mathscr{I}_\theta = ni_\theta = \mathrm{Var}(\dot{\ell}_\theta) = E(-\ddot{\ell}_\theta) \tag{1.1.1}$$

we consider the *a posteriori* observed Fisher information

$$I_n = -\ddot{\ell}_{\hat{\theta}_n}. \tag{1.1.2}$$

Let A_n denote an ancillary statistic. For those families for which exactly ancillary statistics do not exist, we let A_n denote an approximately ancillary statistic. Examples of such statistics are asymptotically ancillary statistics for curved exponential families (Hinkley, 1980) and locally ancillary statistics for families with two-dimensional sufficient statistics (Cox, 1980). We have the following asymptotic results for likelihood statistics conditioned on ancillary statistics (Efron and Hinkley, 1978; Hinkley, 1980; Cox, 1980):

$$\lim_{n\to\infty} P\{I_n(\hat{\theta}_n - \theta)^2 \le c \mid A_n\} = P(\chi_1^2 \le c) + O_p(n^{-1}) \tag{1.1.3}$$

$$\lim_{n\to\infty} P\{2(\ell_{\hat{\theta}_n} - \ell_\theta) \le c \mid A_n\} = P(\chi_1^2 \le c) + O_p(n^{-1}). \tag{1.1.4}$$

Received January 1982; revised September 1982.

AMS 1970 *subject classifications.* Primary 62L05; secondary 62F12.

Key words and phrases. Ancillary, asymptotics, conditional inference, curved exponential family, likelihood ratio, location family, normal circle distribution.

Further, it is *not* the case that

$$\lim_{n\to\infty} P\{\mathscr{I}_{\hat{\theta}}(\hat{\theta}-\theta)^2 \le c \mid A_n\} = P(\chi_1^2 \le c) + O_p(n^{-1}).$$

The rate of convergence is slower. Although, as is well known, $\mathscr{I}_{\hat{\theta}}^{-1/2}$ is the appropriate scale factor in unconditional likelihood inference, it is the wrong conditional scale factor. Because of (1.1.3), we refer to I_n as the true precision of $\hat{\theta}$, in contrast to the expected precision $\mathscr{I}_\theta$.

These two precisions can be quite different and this difference has potential implications for experimental design. Suppose that the model is parameterized in such a way that $i_\theta \equiv i$, a constant. For large n, the difference between I_n and $\mathscr{I}_\theta$ can be quantified by the result that (Efron and Hinkley, 1978)

$$\frac{I_n - ni}{i\sqrt{n}} \sim N(0, \gamma_\theta^2), \tag{1.1.5}$$

where γ_θ^2 is the statistical curvature of $f_\theta(x)$. (See Efron, 1975, for a discussion of statistical curvature.) For example, if $f_\theta(x)$ is the Cauchy density with location parameter θ, for which $i = ½$ and $\gamma_\theta^2 = 5/2$, we find that in approximately 10% of samples, $I_n(x) \le 4.5$ for $n = 20$ although $\mathscr{I} = 10$. Choice of a fixed sample size n to obtain a specified mean position (ni) may yield an experiment with a much lower true precision. A direct way to achieve a given precision for the maximum likelihood estimate is to use a sequential design, sampling sequentially until the observed information reaches a target level.

In this paper, we examine the properties of such a sequential sampling scheme, with particular reference to location families and to curved exponential families. We imagine an infinite sequence $X_1, X_2, \cdots$ of iid random variables with associated sequences $\{\hat{\theta}_n : n = 1, \cdots\}$ and $\{I_n : n = 1, 2, \cdots\}$. Sampling terminates at sample size N such that

$$N = N_{I^*} = \inf\{n : I_n \ge I^*\}, \tag{1.1.6}$$

where I^* specifies the desired precision. For a particular realization of the sequential experiment, we wish to know how to set confidence limits for θ, appropriately conditioned. To summarize what follows, it appears that results (1.1.3) and (1.1.4) continue to hold for large I^* for location families and curved exponential families parameterized so that $i_\theta = i$. Thus, conditional inference proceeds as if n were fixed; the sampling rule is irrelevant to the interpretation of the data.

The proposed sampling scheme is similar to one discussed by Anscombe (1952). He considered a sequential sampling scheme designed to achieve a specified level of unconditional precision and showed that fixed sample size asymptotic results continue to hold in the sequential schemes. Lindley (1957) suggested a Bayesian approach to the problem of sequential sampling in the context of the estimation of the unknown proportion in a binomial population. He compared different sampling schemes designed to achieve a specified level of Shannon information in the posterior distribution for various parameterizations, and showed that they are asymptotically equivalent to sampling schemes designed to achieve a specified level of Fisher observed information. Chow and Robbins (1965) considered the problem of finding a confidence interval of prescribed width and coverage probability for the mean of a population with unknown variance. They showed that a sampling scheme similar to the one described here, in which one sampled data until the estimate of the precision of the mean exceeded a prescribed level, was asymptotically consistent and efficient.

Section 2 examines the theory of conditional sequential estimation. Section 2.1 proves that sequential versions of (1.1.3) and (1.1.4) hold for local families. Section 2.2 considers two-dimensional curved exponential families. It shows that the fixed sample size approximately ancillary statistic is also approximately ancillary under sequential sampling and conjectures that (1.1.3) and (1.1.4) hold for sequential samples. Section 2.3 proves the asymptotic normality of the sequential sample size. The results of two small Monte Carlo simulations, a Cauchy location family and a bivariate normal family with unknown correlation coefficient, are presented in Section 3. Section 4 gives a brief summary.

2. Theory of conditional sequential estimation.

2.1. *Location families.* In this section, we show that (1.1.3) and (1.1.4) hold in the sequential case for location families, families in which $f_\theta(x) = f_0(x - \theta)$. In location families, $i_\theta \equiv i$ and $\gamma_\theta^2 \equiv \gamma^2$. We begin with a brief view of the fixed sample size results.

Let $s_n = (X_1, \cdots, X_n)$ be a random sample from $f_\theta(x)$. If we let $d_n = x_n - x_1$, the labeled distance from the first observation, then we have the following exactly ancillary statistic:

$$a_n = (d_2, d_3, \cdots, d_n). \tag{2.1.1}$$

Fisher (1934) proved the equivalence of the likelihood function when reflected about its maximum and suitably normalized to the conditional density of the pivotal quantity, $\hat{\theta}_n - \theta$. Let $t_n = \hat{\theta}_n - \theta$. We have

$$f(t_n \mid a_n) = \exp\{\ell_{\hat{\theta}(s_n) - t_n}(s_n)\} \Big/ \int \exp\{\ell_u(s_n)\}\, du. \tag{2.1.2}$$

(This equation concisely summarizes the following manipulation: In location families, the likelihood function for any sample s_n depends on θ solely through the quantity $\theta - \hat{\theta}(s_n)$, where $\hat{\theta}(s_n)$ is the MLE of the sample. If one substitutes the argument $\hat{\theta} - \theta$ in which $\hat{\theta}$ is considered as a random variable for the argument $\theta - \hat{\theta}(s_n)$ in the likelihood function and normalizes the function so that its integral over R is 1, the result is the conditional density of $\hat{\theta} - \theta$.)

We note that in (2.1.2) we can substitute for a_n that function of a_n which corresponds to the maximal invariant after minimal sufficient reduction. Efron and Hinkley (1978) applied the results of Walker (1969) on the asymptotic normality of the likelihood function to prove (1.1.3) and (1.1.4) from (2.1.2).

We now examine the sequential sampling scheme (1.1.6). Considerations of invariance show that the sequence $(d_2, d_3, \cdots, d_n, \cdots)$ is the maximal ancillary.

Any function of this sequence is ancillary. Specifically, the following are ancillary:

$$\text{(i) } (I_1, I_2, \cdots, I_N; N), \quad \text{(ii) } I_N, \quad \text{and} \quad \text{(iii) } a_N = (d_2, \cdots, d_N). \tag{2.1.3}$$

THEOREM 2.1.1. *Under the regularity conditions listed below, as $I^* \to \infty$ in the sequential sampling scheme* (1.1.6) *where the X_i's are sampled from a location family, we have*

$$P\{I_N(\hat{\theta}_N - \theta)^2 \le c \mid A_N = a_N\} = P(\chi_1^2 \le c) + O_p(I^{*-1}) \tag{2.1.4}$$

and

$$P\{2(\ell_{\hat{\theta}} - \ell_\theta) \le c \mid A_N = a_N\} = P(\chi_1^2 \le c) + O_p(I^{*-1}). \tag{2.1.5}$$

CONDITIONS.

2.1.1(i). The following three integrals are finite almost surely (a.s.):

$$\text{(a)} \quad \int_{-\infty}^{\infty} \exp(\ell_{\hat{\theta}-t} - \ell_{\hat{\theta}})\, dt,$$

$$\text{(b)} \quad \int_{-\infty}^{\infty} \exp\{(1-2s)(\ell_{\hat{\theta}-t} - \ell_{\hat{\theta}})\}\, dt \quad \text{for} \quad s < \frac{1}{2},$$

(c) $$\int_{-\infty}^{\infty} \exp\{sI_N t^2(\ell_{\hat\theta-t} - \ell_{\hat\theta})\}\, dt \quad \text{for} \quad s < \frac{1}{2}.$$

2.1.1(ii). The first four derivatives of log $f_\theta(x)$ with respect to θ exists a.s. in X for all θ and have finite expectations under the true value of θ. The second derivative has a strictly negative expected value under the true value of θ. The fourth derivative is continuous everywhere in θ a.s. in X.

2.1.1(iii). Let $\ell_\theta^j = \partial\ell_\theta^j/\partial\theta^j\,|_{\theta=\hat\theta}$. We have $\ell_{\hat\theta}^j/I_N = O_p(1)$ for $j = 3, 4, \cdots$, and $I_N/I^* = O_p(1)$.

2.1.1(iv). For arbitrarily small $\delta > 0$, there exists $c_\delta > 0$ such that

$$\lim_{I^*\to\infty} P\{\sup_{|t|>\delta} I^{*-1}(\ell_{\hat\theta-t} - \ell_{\hat\theta}) < -c_\delta\} = 1.$$

These conditions have been quite easy to check in the examples examined by the author.

PROOF. Because the ancillary statistic a_N in the sequential sampling scheme is identical to the ancillary a_n in a fixed sized sampling scheme which results in the same sample and because the sequence $I_2, \cdots I_N$ which determines the stopping time N is a function of a_N only, conditioning on the ancillary statistic in either scheme gives the identical reference set in $\mathbb{R}^n$. Therefore, we can apply Fisher's (1934) result to find $f(t_N \mid a_N) = \exp\{\ell_{\hat\theta}(s_N)_{-t_N}(s_N)\}/\int \exp\{\ell_u(s_N)\}\, du$. We now need merely utilize the proof in Efron and Hinkley (1978) for (1.1.3) and (1.1.4) which also starts with Fisher's 1934 result, modifying it slightly to take account of the different sampling scheme. Condition 2.1.1(iv) guarantees that the likelihood function will behave appropriately as I^* increases. Efron and Hinkley (1978) use a similar condition, replacing I^* by n. For the technical details, see Grambsch (1980).

In many location families, the minimal sufficient statistic provides substantial dimensionality reduction. Suppose that a one-to-one function transforms the minimal sufficient statistic into the pair of statistics, the MLE and the maximal invariant after minimal sufficient reduction. Let b_n denote the maximal invariant after minimal sufficient reduction in the fixed sample size case. Because

$$f(t_n \mid (b_2, \cdots, b_n)) = f(t_n \mid b_n) = \exp\{\ell_{\hat\theta(s_n)-tn}(s_n)\}/\exp\{\ell_\mu(s_n)\}\, du,$$

reasoning similar to that used in the proof of Theorem 2.1.1 can be used to prove the following corollary.

COROLLARY 2.1.1. *Under the regularity conditions for Theorem* 2.1.1,

$$P\{I_N(\hat\theta_N - \theta)^2 \le c \mid B_N = b_n\} = P(\chi_1^2 \le c) + O_p(I^{*-1}). \tag{2.1.6}$$

$$P\{2(\ell_{\hat\theta} - \ell_\theta) \le c \mid B_N = b_n\} = P(\chi_1^2 \le c) + O_p(I^{*-1}). \tag{2.1.7}$$

When the minimal sufficient statistic is two-dimensional, the maximal invariant after minimal sufficient reduction is the observed information. Example 2.1.1 discusses such a location family. Section 3 shows that a transformation of the observed information functions as the asymptotic ancillary statistic in curved exponential families with two dimensional sufficient statistics.

It is noteworthy that the size of the sequential sample is irrelevant to the applicability of Theorem 2.1.1. Anscombe's theorem on unconditional sequential inference, Theorem 2.2.1, mentioned in Section 1, requires a large sequential sample in order that the asymptotic fixed sample size distribution of the estimator provide a good approximation to the sequential distribution. However, Theorem 2.1.1 does not require large sample size; it simply requires large I^*. In this respect, the sampling scheme is similar to the sampling scheme designed to achieve fixed width confidence intervals (Chow and Robbins, 1965) mentioned in the introduction. The applicability of the fixed-width confidence interval requires large achieved sample precision, irrespective of sample size. For sufficiently large I^*, the χ_1^2 distribution may be a good approximation to the conditional distribution of $2(\ell_{\hat{\theta}} - \ell_\theta)$ and $I_N(\hat{\theta} - \theta)^2$ given the appropriate ancillary for sample sizes as small as one. The following example provides an illustration.

EXAMPLE 2.1.1. Consider the normal circle example discussed by Efron and Hinkley (1978). A sample point consists of a vector sampled from a bivariate normal distribution with covariance matrix the identity whose mean vector lies on a circle of known radius ρ centered at the origin. That is,

$$E(X^T) = \rho(\cos\theta, \sin\theta). \tag{2.1.8}$$

Note that given n independent bivariate observations $X_1, \cdots, X_n$, the minimal sufficient statistic $s = \sum x_i/\sqrt{n}$ satisfies (2.1.8) with $\rho_n = \rho\sqrt{n}$ in place of ρ. Let s have polar coordinates $(\hat{\theta}, r_n\rho_n)$. The joint density of $(\hat{\theta}, r_n^2\rho_n^2)$ is

$$f(\hat{\theta}, r_n^2\rho_n^2) = \frac{1}{4\pi}\exp\left\{-\frac{1}{2}\rho_n^2(r_n^2+1)\right\}\exp\{\rho_n^2 r_n\cos(\hat{\theta}-\theta)\},$$

from which it can be seen that the density is of location form with $\hat{\theta}$ as the maximum likelihood estimate of θ. Considerations of invariance show that r_n is ancillary. The conditional distribution of $\hat{\theta}$ given r_n is a von Mises distribution,

$$f(\hat{\theta} \mid r_n) = c^{-1}\exp\{\rho_n^2 r_n\cos(\hat{\theta}-\theta)\}, \tag{2.1.9}$$

where $c = 2\pi I_0(\rho_n^2 r_n)$ is standard Bessel function notation. We note that $I_n = \rho_n^2 r_n$, so the observed information is ancillary. Therefore, with n replaced by N, (2.1.9) applies in the sequential scheme (1.1.6). It is clear that $I_1 = \rho_1^2 r_1$ may be arbitrarily large with nonzero, albeit small, probability. The application of standard results for the von Mises distribution to (2.1.9) shows that conditional on r_1, $(\hat{\theta}_1 - \theta)\rho_1\sqrt{r_1}$ is asymptotically standard normal for large values of $\rho_1^2 r_1$. From this fact, we can easily derive (2.1.4) and (2.1.5) for samples of size 1. (The precise order of the error term would require a fuller analysis.) Of course, Theorem 2.1.1 could be applied directly to this example as well.

2.2 *Ancillary statistics in a curved exponential family.* In this section we consider curved exponential families in two dimensions. The density takes the following form

$$f_\theta(x) = g(x)\exp\{\eta'(\theta)s(x) + \psi(\theta)\}.$$

Here $s(x)$ is a two-dimensional minimal sufficient statistic and η is a two-dimensional parameter vector. We suppose that η is a continuously twice-differentiable function of θ, a scalar parameter. Curved exponential families are discussed in Efron (1975). Example 2.1.1 is an example of a curved exponential family. We suppose that the model is parameterized in such a way that $i_\theta \equiv i$, a constant.

Consider a random sample of size n from a curved exponential family model. In general, an exactly ancillary statistic does not exist. However, Efron and Hinkley (1978) have shown that

$$A_n = \frac{\sqrt{n}}{\gamma_{\hat{\theta}}}\left(\frac{I_n}{ni} - 1\right) \tag{2.2.1}$$

is asymptotically standard normal (cf. (1.1.5)). It functions as an approximate ancillary for two-dimensional curved exponential families and as the dominant component of the $(k-1)$-dimensional ancillary statistic for general k-dimensional curved exponential families. Further, in the parameterization in which $i_\theta \equiv i$, Hinkley (1980) has shown that

Fisher's (1934) result applies approximately. That is,

$$f(\hat{\theta} \mid a_n) = \frac{\exp\{\ell_{\theta}(s_n)\}}{\int \exp(\ell_t(s_n)\}\,dt}\{1 + O(n^{-1})\}.$$

From this one can prove (1.1.3) and (1.1.4) (Hinkley, 1980; Cox, 1980).

We now consider the sequential sampling scheme (1.1.6). A natural sequential version of the fixed sample size ancillary (2.2.1) is

(2.2.2)
$$A_N = \frac{\sqrt{N}}{\gamma_{\hat{\theta}_N}}\left(\frac{I_N}{Ni} - 1\right).$$

We will apply a theorem of Anscombe (1952) to show that A_N is asymptotically standard normal as I^* tends to infinity.

Anscombe's theorem is the following:

Let $\{Y_n, n = 1, 2, \cdots\}$ be an infinite sequence of random variables for which there exist a real number θ, a sequence of positive numbers $\{w_n\}$ and a distribution function G such that

(2.2.3)
$$\lim_{n\to\infty} P\left(\frac{Y_n - \theta}{w_n} \leq x\right) = G(x)$$

at each continuity point of G. Let N_r be a sequence of random stopping times. If the conditions below are met, then

THEOREM 2.2.1 (Anscombe, 1952).

$$\lim_{r\to\infty} P\left(\frac{Y_{N_r} - \theta}{w_{n_r}} \leq x\right) = G(x)$$

at continuity points of G.

CONDITIONS

2.1.1(i). Uniform continuity of $\{Y_n\}$: Given any small positive ε and ξ, there exists a large V and a small positive c such that for any $n > V$,

$$P(|Y_{n'} - Y_n| < \varepsilon w_n \ \forall n' \text{ such that } |n' - n| < cn) > 1 - \xi.$$

2.2.1(ii) Convergence of $\{N_r\}$: Let $\{n_r\}$ be an increasing sequence of positive numbers tending to infinity. We have $P(N_r < \infty) = 1$ for all r and $N_r/n_r \to_P 1$ as $r \to \infty$.

We will apply Theorem 2.2.1 to prove the asymptotic standard normality of A_N. In the notation of that theorem, $Y_n = I_n/n$, $\theta = 1$, $w_n = n^{-1/2} i\gamma_{\theta}$, $n_r = E(N_{I^*})$ and $N_r = N_{I^*}$. We know that (2.2.3) holds by virtue of (1.1.5). In Section 2.3, which discusses the asymptotic distribution of N, we will see that condition 2.2.1(ii) is met. We show that Y_n satisfies 2.2.1(i) by showing that it is asymptotically equivalent to a statistic X_n which satisfies 2.2.1(i).

LEMMA 2.2.2. *Let*

$$\varepsilon_n = n^{-1/2}\left| I_n - \sum_{j=1}^{n}\left[-\ddot{\ell}_{\theta}(x_j) + \frac{E\{\dot{\ell}_{\theta}(x_1)\ddot{\ell}_{\theta}(x_1)\}}{i}\dot{\ell}_{\theta}(x_j)\right]\right|.$$

Under the regularity conditions 2.2.2(i)–(iii) *below, in a fixed sample size sampling scheme*

$$\varepsilon_n \to 0 \quad \text{a.s. } and \quad \varepsilon_n = O_p(n^{-1/2}).$$

CONDITIONS.

2.2.2(i). The first four derivatives of $\log f_\theta(x)$ with respect to θ exist a.s. in X in an open neighborhood including the true value of θ.

2.2.2(ii). For all θ in an open neighborhood including the true value,

$$|\dot{\ell}_\theta^4(x_1)| < M(x) \quad \text{where} \quad E\{M(X)\} < K < \infty.$$

2.2.2(iii). $E\{-\dot{\ell}_\theta(x_1)\ddot{\ell}_\theta(x_1)\} = E\{\dddot{\ell}_\theta(x_1)\}$.

These conditions are met in curved exponential families (Efron and Hinkley, 1978).

PROOF. The result follows from a Taylor series expansion of I_n about the true value of θ and the use of the strong law of large numbers. Let

$$\bar{Y}_n = n^{-1}\left[\sum_{j=1}^n - \ddot{\ell}_\theta(x_j) + \frac{E\{\dot{\ell}_\theta(x_1)\ddot{\ell}_\theta(x_1)\}}{i}\sum_{j=1}^n \dot{\ell}_\theta(x_j)\right];$$

$\bar{Y}_n$ is a sample average of iid random variables with mean i and variance $i^2\gamma_\theta^2$. Thus, $w_n = n^{-1/2}i\gamma_\theta$ is the scale factor for $\bar{Y}_n$. In the same paper, Anscombe showed that sample averages and statistics asymptotically equivalent to sample averages satisfy 2.2.1(i). By Lemma 2.2.2, I_n/n satisfies 2.2.1(i). Therefore, we have shown that

$$\frac{\sqrt{E(N)}}{\gamma_\theta}\left(\frac{I_N}{Ni} - 1\right) \to_{\mathscr{L}} N(0, 1) \quad \text{as} \quad \mathscr{I}^* \to \infty. \tag{2.2.4}$$

But, as will be shown in Section 2.3,

$$\frac{\sqrt{N}}{\sqrt{E(N)}} \to_P 1. \tag{2.2.5}$$

Further, it is an easy application of Anscombe's Theorem 2.2.1 to show that $\hat{\theta}_N \to_P \theta$ and therefore, if γ_θ is a continuous function of θ,

$$\gamma_\theta/\gamma_{\hat{\theta}_N} \to_P 1. \tag{2.2.6}$$

Thus, we have established

THEOREM 2.2.2. *Suppose Conditions* 2.2.2(i)–(iii) *are met and, in addition,* γ_θ *is a continuous function of* θ. *Then*

$$\frac{\sqrt{N}}{\gamma_{\hat{\theta}_N}}\left(\frac{I_N}{Ni} - 1\right) \to_{\mathscr{L}} N(0, 1) \quad \text{as} \quad I^* \to \infty.$$

This result, showing that A_N given by (2.2.2) functions as an approximate ancillary in the sequential sampling scheme, leads to the conjecture that for two-dimensional curved exponential families with $i_\theta \equiv i$, the following asymptotic results hold:

$$P\{I_N(\hat{\theta}_N - \theta)^2 \le c \mid A_n = a\} \to P(\chi_1^2 \le c) \quad \text{as} \quad I^* \to \infty \tag{2.2.7}$$

and

$$P\{2(\ell_{\hat{\theta}_N} - \ell_\theta) \le c \mid A_n = a\} \to P(\chi_1^2 \le c) \quad \text{as} \quad I^* \to \infty. \tag{2.2.8}$$

In Section 3, we present evidence for this conjecture by means of the results of a Monte Carlo simulation.

2.3 *Asymptotic distribution of N.* In this section we show that the sequential size N has an asymptotic normal distribution when suitably standardized.

THEOREM 2.3.1 *Under the sequential sampling scheme* (1.1.6) *and given the regularity conditions* 2.2.2(i)–(iii) *for Lemma* 2.2.2,

$$\frac{N - (I^*/i)}{\gamma_\theta \sqrt{(I^*/i)}} \rightarrow_{\mathscr{L}} N(0, 1) \quad \text{as} \quad I^* \rightarrow \infty.$$

PROOF. The author gratefully acknowledges the comments of the Associate Editor in suggesting a simpler proof of this theorem than that originally presented. In fact, the theorem follows by standard methods from the following two results: (i) a theorem (see Theorem 2, Section 9.4 of Chow and Teicher, 1978) on the asymptotic normality of stopping times for sequences of sums of iid random variables, and (ii) Lemma 2.2.2, which shows that I_n when suitably normalized is asymptotically equivalent to a sum of iid random variable in the strong sense of convergence with probability one.

3. Monte Carlo simulations. In this section we present the results from two small Monte Carlo simulations, one for a location family and one for a two-dimensional curved exponential family.

The location family was the Cauchy translation family with pdf

$$f_\theta(x) = \pi^{-1}\{1 + (x - \theta)^2\}^{-1}.$$

The Univesity of Minnesota CDC Cyber-172 was used to generate 1000 sequential samples, with $\theta = 0$ and $I^* = 10.0$. For each sample, the sample size N and the MLE, $\hat{\theta}_N = \hat{\theta}$, were obtained and three statistics were computed. Two of them were the two asymptotic χ_1^2 statistics of the theory, $2(\ell_{\hat{\theta}} - \ell_\theta)$ and $I_N(\hat{\theta} - \theta)^2$. For comparison, the asymptotic χ_1^2 for unconditional inference for the fixed sample size sampling scheme, $\mathscr{I}_{\hat{\theta}}(\hat{\theta} - \theta)^2$ was obtained. In this case, $\mathscr{I}_{\hat{\theta}} = N/2$. We examined the distribution of the three statistics conditional on N, which was shown to be ancillary in Section 2.1. The samples were divided into nine roughly equal groups on the basis of N. For each group the proportion of values of each statistic exceeding various upper-tail values of χ_1^2 distribution was obtained. Table 3.1 shows the results for the 95th percentile. The results for the 90th and 99th percentiles are similar. The standard error of each proportion is approximately 0.02. The behavior of these statistics is in accordance with the theory. The proportion of values above 3.84 varies randomly with N for the two statistics $2(\ell_{\hat{\theta}} - \ell_\theta)$ and $I_N(\hat{\theta} - \theta)^2$. For $\mathscr{I}_{\hat{\theta}}(\hat{\theta} - \theta)^2$ the proportion above 3.84 increases with N because $\mathscr{I}_{\hat{\theta}} = N/2$. The proportion above 3.84 is closer to the theoretical value of 0.05 for the statistic $2(\ell_{\hat{\theta}} - \ell_\theta)$ than for $I_N(\hat{\theta} - \theta)^2$. The improved behavior of the likelihood ratio statistic is also found in the fixed sample size sampling scheme (Efron and Hinkley, 1978) and is in accordance with the higher order expansion of the error terms in the asymptotic conditional distribution of these statistics (Grambsch, 1980).

TABLE 3.1
Proportions of Simulation Statistics Exceeding 3.84, *Grouped by N, in Cauchy example*

N	Frequency	Statistics		
		$2(\ell_{\hat{\theta}} - \ell_\theta)$	$I(\hat{\theta} - \theta)^2$	$I_{\hat{\theta}}(\hat{\theta} - \theta)^2$
6–11	96	.042	.052	0
12–13	95	.032	.042	.021
14–15	118	.093	.110	.059
16–17	128	.055	.055	.047
18–19	133	.046	.053	.030
20–21	115	.026	.026	.026
22–24	121	.066	.099	.099
25–28	105	.048	.086	.162
29–52	89	.034	.034	.090
average proportion		.050	.063	.060

TABLE 3.2

Empirical conditional distribution of $T = 2(\ell_{\hat{\phi}_N} - \ell_\phi)$ given a_N for $I^* = 30$ in correlation example. For each decile interval of a_N, entries are observed frequencies (in 100 samples) with which T falls in stated χ_1^2 percentile range.

Theoretical χ_1^2 percentile range	**a_N decile**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0–10	11	13	14	11	6	19	11	13	10	11
10–20	11	11	5	18	9	10	11	14	10	12
20–30	7	10	11	11	8	13	7	9	10	13
30–40	7	7	15	6	8	6	11	3	5	12
40–50	11	9	10	8	10	11	7	10	8	8
50–60	9	11	1	9	13	8	12	4	14	5
60–70	12	10	11	10	17	11	10	12	7	10
70–80	6	15	3	9	8	6	5	11	8	9
80–90	10	5	9	8	11	6	13	11	11	11
90–95	6	6	6	5	6	5	7	7	12	5
95–100	10	3	5	5	4	5	6	6	5	4

TABLE 3.3

Empirical conditional distribution of $T = I_N(\hat{\phi}_N - \phi)^2$ given a_N for $I^* = 30$ in correlation example. (Entries as defined for Table 3.2.)

Theoretical χ_1^2 percentile range	**a_N decile**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0–10	11	13	14	11	6	19	11	13	10	11
10–20	11	11	5	18	9	10	11	14	10	12
20–30	7	9	11	11	8	13	7	9	10	13
30–40	6	8	15	6	7	6	11	3	5	12
40–50	12	8	10	8	11	11	7	10	8	8
50–60	9	11	10	9	11	8	12	4	14	5
60–70	12	10	10	9	19	10	10	11	7	10
70–80	5	15	4	10	6	7	5	12	8	10
80–90	10	5	9	8	13	6	12	10	11	10
90–95	5	6	7	4	3	4	7	7	12	5
95–100	13	4	5	6	7	6	7	7	5	4

The second Monte Carlo simulation is a two-dimensional curved exponential family. We illustrate the conjectures of Section 2, (2.2.7) and (2.2.8) by means of the normal correlation example discussed in Efron and Hinkley (1978). Let $\{(x_{1i}, x_{2i}), i = 1, 2, \cdots\}$ be a sequence of iid pairs sampled from a bivariate normal density with mean vector zero, unit variances, and correlation θ. The information in this parameterization is not constant. We reparameterized to $\phi = \sqrt{2}\,\tanh^{-1}(\eta_\theta\sqrt{2}) - \tanh^{-1}(\eta_\theta)$ where $\eta_\theta = \theta(1 + \theta^2)^{-1/2}$. We note that $i_\phi \equiv 1$.

Several very small Monte Carlo simulations were done on the University of Minnesota CDC Cyber-172 using various values of I^* and ϕ and the results were in accord with the conjecture. We present the results for $I^* = 30.0$ and $\phi = \theta = 0.0$. One thousand sequential samples were generated. The conjecture that conditional on a_N, $I_N(\hat{\phi}_N - \phi)^2$ and $2(\ell_{\hat{\phi}_N} - \ell_\phi)$ would be distributed approximately as χ_1^2 random variables was examined by obtaining empirical conditional distributions for the two statistics. The empirical distribution for the ancillary statistic was divided exactly into deciles. A grouped empirical frequency distribution was obtained for each of the two observed likelihood statistics for the 100 sequential samples whose ancillary fell into each decile. The groups corresponded to the deciles of a χ_1^2 distribution, with the exception that the 10th decile was divided at the 95th percentile to examine tail behavior. The results are shown in Tables 3.2 and 3.3.

The frequencies expected under a χ_1^2 distribution would be 10 in each of the first 9 rows and five in each of the last two. The empirical data fit the χ_1^2 distribution quite well in both tables.

For each table, we computed the Pearson goodness-of-fit statistic $x^2 = \sum(O - E)^2/E$ which is asymptotically distributed as a χ_{100}^2. We found $x^2 = 98.10$ for $2(\ell_{\hat{\phi}_N} - \ell_\phi)$ and $x^2 = 111.70$ for $I_N(\hat{\phi} - \phi)^2$. There is good reason to be satisfied with the hypothesis that the statistics have conditional χ_1^2 distributions given the ancillary statistic. These two Monte Carlo simulations show that the conditional distribution of the two likelihood statistics is approximated quite well by a χ_1^2 even for moderate values of I^*.

4. Summary and conclusions. In this paper, we have examined a sampling scheme in which observations are taken sequentially until the precision of $\hat{\theta}$ reaches a desired level, and have shown that asymptotic conditional inference based on likelihood statistics is identical to that in the fixed sample size experiment. More theoretical work is needed for the curved exponential family and extensions of it. Useful applications should be examined.

Acknowledgments. The author wishes to thank David Hinkley for many useful discussions on this topic.

REFERENCES

ANSCOMBE, F. J. (1952). Large sample theory of sequential estimation. *Proc. Camb. Phil. Soc.* **48** 600–607.

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.

CHOW, Y. S. and TEICHER, H. (1978). *Probability Theory.* Springer-Verlag, New York.

COX, D. R. (1980). Local ancillarity. *Biometrika* **67** 279–286.

EFRON, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency) (with Discussion). *Ann. Statist.* **3** 1189–1242.

EFRON, B. and HINKLEY, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: observed vs. expected Fisher information (with Discussion). *Biometrika* **65** 457–487.

FISHER, R. A. (1934). Two new properties of mathematical likelihood. *Proc. Roy. Statist. Soc. Ser. A* **144** 285–307.

GRAMBSCH, P. L. (1980). Conditional likelihood inference. Ph.D. Thesis, University of Minnesota.

HINKLEY, D. V. (1980). Likelihood as approximate pivotal distribution. *Biometrika* **67** 287–292.

LINDLEY, D. V. (1957). Binomial sampling schemes and the concept of information. *Biometrika* **44** 179–186.

WALKER, A. M. (1969). On the asymptotic behavior of posterior distributions. *J. Roy. Statist. Soc. Ser. B* **31** 259–270.

BELL LABORATORIES
600 MOUNTAIN AVENUE
MURRAY HILL, NEW JERSEY 07974