The Annals of Statistics
1977, Vol. 5, No. 1, 191–196

OPTIMAL ALLOCATION OF OBSERVATIONS IN INVERSE LINEAR REGRESSION

BY S. K. PERNG AND Y. L. TONG

Kansas State University and University of Nebraska

Consider the problem of estimating x under the inverse linear regression model

$$Y_i = \alpha + \beta x_i + \varepsilon_i\,, \qquad Z_j = \alpha + \beta x + \varepsilon_j'$$

for $i = 1, \cdots, n, \cdots, j = 1, \cdots, m, \cdots$, where $\{\varepsilon_i\}$, $\{\varepsilon_j'\}$ are two sequences of i.i.d. rv's with 0 means and finite variances, $\{x_i\}$ is a sequence of known constants and α, β, x are unknown parameters. For fixed $T = m + n$, this paper considers a sequential procedure for the optimal allocation of m and n. It is shown that, as $T \to \infty$, the procedure is asymptotically optimal.

1. Introduction. Consider the following model of the inverse linear regression problem:

$$\begin{aligned} Y_i &= \alpha + \beta x_i + \varepsilon_i & i &= 1, 2, \cdots, n, \cdots \\ Z_j &= \alpha + \beta x + \varepsilon_j' & j &= 1, 2, \cdots, m, \cdots \end{aligned} \tag{1.1}$$

where $\{\varepsilon_i\}$, $\{\varepsilon_j'\}$ are two sequences of i.i.d. random variables with means 0 and finite unknown variances $\sigma_1^2 > 0$, $\sigma_2^2 > 0$, respectively; $\{x_i\}$ is a sequence of known constants, and α, β and x are unknown. Under the assumptions that the random variables are normally distributed with $\sigma_1^2 = \sigma_2^2$ and that n, m are predetermined, the point and interval estimations of x have been studied previously (e.g., [6], [7]). In this paper we consider, under more general conditions, the optimal allocation of n (and m) for the interval estimation of x so that the probability of coverage is maximized when the total number of observations $T = n + m$ is fixed and is large.

In Section 2 the coverage probability function (of the ratio $\theta = \lim_{T\to\infty}(n/T)$) is investigated. Bounds on the optimal value of θ are given, and a sequential procedure is considered in Section 3 so that the observations may be allocated, one at a time, for observing either a Y_i or Z_j. It is shown that this procedure is asymptotically optimal as $T \to \infty$. Monte Carlo results are given in Section 4.

2. Asymptotic theory and the coverage probability function. For $n = 1, 2, \cdots$ let

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^n x_i\,, \qquad S_n^2 = \sum_{i=1}^n (x_i - \bar{x}_n)^2\,. \tag{2.1}$$

We shall restrict our attention to those $\{x_i\}$ sequences satisfying, as $n \to \infty$, (a)

Received June 1974; revised October 1975.
AMS 1970 *subject classifications.* Primary 62J05; Secondary 62L12.
Key words and phrases. Inverse linear regression, allocation of observations, sequential methods.

$(1/n)\max_{1\leq i\leq n} x_i^2 \to 0$, (b) there exists two real numbers μ and $c > \mu^2$ such that $\bar{x}_n \to \mu$ and $(1/n)\sum_{i=1}^n x_i^2 \to c$. Since the problem remains unchanged when the x_i's are replaced by the values obtained through a linear transformation, without loss of generality it is assumed that $\mu = 0$ and $c = 1$.

Now for observed Y_i, Z_j, $i = 1, \cdots, n$, $j = 1, \cdots, m$ consider the estimator

$$\hat{x} = (\bar{Z} - \hat{\alpha})/\hat{\beta} \tag{2.2}$$

where

$$(\hat{\alpha}, \hat{\beta})' = (\mathbf{X}_n\mathbf{X}_n')^{-1}\mathbf{X}_n\mathbf{Y}_n, \tag{2.3}$$

$\bar{Z} = (1/m)\sum_{j=1}^m Z_j$, $\mathbf{Y}_n' = (Y_1, \cdots, Y_n)$ and

$$\mathbf{X}_n = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{pmatrix}. \tag{2.4}$$

Since $\hat{\alpha}$, $\hat{\beta}$ are unbiased estimators of α, β it follows that

$$\frac{n^{\frac{1}{2}}}{\sigma_1}(\hat{\alpha} - \alpha, \hat{\beta} - \beta)' = n^{\frac{1}{2}}(\mathbf{X}_n\mathbf{X}_n')^{-1}\mathbf{X}_n\mathbf{V}_n$$

where $\mathbf{V}_n' = (V_1, \cdots, V_n)$, $V_i = (1/\sigma_1)(Y_i - EY_i)$, $i = 1, \cdots, n$ and $V_1, \cdots, V_n, \cdots$ is a sequence of i.i.d. random variables with zero mean and unit variance. It follows from a theorem in [5] (page 153) that

$$\frac{1}{n^{\frac{1}{2}}}\mathbf{X}_n\mathbf{V}_n \to_d N(0, I) \qquad \text{as} \quad n \to \infty,$$

where I is the (2×2) identity matrix. Let $T = m + n$ denote the total number of observations available to the experimenter and let $\theta_{m,n} = (n/T)$, $\lim_{T\to\infty}\theta_{m,n} = \theta \in (0, 1)$, $\delta = \sigma_1^2/\sigma_2^2$. We have thus obtained

THEOREM 1. *For every* $\theta \in (0, 1)$,

$$T^{\frac{1}{2}}(\hat{\alpha} - \alpha, \hat{\beta} - \beta, \bar{Y} - \alpha - \beta x)' \to_d N(\mathbf{0}, \sigma_1^2 D)$$

as $T \to \infty$, *where* D *is a diagonal matrix with elements* $d_{11} = d_{22} = 1/\theta$ *and* $d_{33} = \delta/(1 - \theta)$.

Now consider a sequence of positive real numbers $\{d_T\}$, denote the confidence interval for x by $(\hat{x} - d_T, \hat{x} + d_T)$. Then the probability of coverage is $P[|\hat{x} - x| \leqq d_T]$. For $\theta \in (0, 1)$ let $\tau = \tau(x, \delta, \theta)$ be

$$\tau = (1 + x^2)/\theta + \delta/(1 - \theta). \tag{2.5}$$

THEOREM 2. *If* $\beta \neq 0$ *and the sequence* $\{d_T\}$ *satisfies* $T^{\frac{1}{2}} d_T \to a$ *where* a *is either* ∞ *or a finite real number, then*

$$\lim_{T\to\infty} P[|\hat{x} - x| \leqq d_T] = \Phi(a|\beta|/(\sigma_1\tau^{\frac{1}{2}})) - \Phi(-a|\beta|/(\sigma_1\tau^{\frac{1}{2}})) = g(\theta), \quad \textit{say}, \tag{2.6}$$

where Φ *is the standard normal* cdf.

PROOF. Applying Theorem 1 and a theorem in [1] (page 76) yields

$$T^{\frac{1}{2}}(\hat{x} - x) \to_d N(0, \tau(\sigma_1/\beta)^2) .$$

If $a = \infty$, then (2.6) is obvious. Otherwise, for fixed $\varepsilon > 0$ let T_0 be such that

$$|P[|\beta|T^{\frac{1}{2}}(\hat{x} - x)/(\sigma_1\tau^{\frac{1}{2}}) \leqq z] - \Phi(z)| < \varepsilon \qquad \text{uniformly in} \quad z ,$$

and

$$|\{\Phi(|\beta|T^{\frac{1}{2}}\, d_T/(\sigma_1\tau^{\frac{1}{2}})) - \Phi(-|\beta|T^{\frac{1}{2}}\, d_T/(\sigma_1\tau^{\frac{1}{2}}))\} - g(\theta)| < \varepsilon$$

hold simultaneously whenever $T > T_0$ (the existence of T_0 is assured by the continuity of Φ and the uniform convergence in distribution to Φ). Then

$$|P[|\hat{x} - x| \leqq d_T] - g(\theta)| \leqq 3\varepsilon$$

holds whenever $T > T_0$. This completes the proof of the theorem.

Let θ_0 satisfy $g(\theta_0) = \sup_\theta g(\theta)$. This θ_0 is the approximate solution for the problem of optimal allocation of observations when T is large, and it can be obtained by minimizing $\tau(x, \delta, \theta)$. It is easily seen that

$$\theta_0 = (1 + x^2)^{\frac{1}{2}}/((1 + x^2)^{\frac{1}{2}} + \delta^{\frac{1}{2}}) \tag{2.7}$$

or equivalently, $\theta_0/(1 - \theta_0) = ((1 + x^?)/\delta)^{\frac{1}{2}}$; which gives $\tau_0 = \tau(x, \delta, \theta_0) = ((1 + x^2)^{\frac{1}{2}} + \delta^{\frac{1}{2}})^2$ and

$$g(\theta_0) = \Phi(a|\beta|/\{\sigma_1((1 + x^2)^{\frac{1}{2}} + \delta^{\frac{1}{2}})\}) - \Phi(-a|\beta|/\{\sigma_1((1 + x^2)^{\frac{1}{2}} + \delta^{\frac{1}{2}})\}) . \tag{2.8}$$

3. A sequential procedure and its asymptotically optimal properties. Since the optimal ratio of allocation θ_0 given by (2.7) depends on both δ and x where x is the parameter one wishes to estimate, the problem of optimal allocation of observations cannot be solved when n and m are predetermined. In the following a sequential procedure is proposed for this purpose and its asymptotically optimal properties (as $T \to \infty$) are investigated. The idea under this procedure is to allocate the fixed number of T observations sequentially so that at each step estimators of δ and x are calculated, a decision is then made on where the next observation should be taken from.

PROCEDURE R. (a) For arbitrary but fixed $n_0 \geqq 3$, $m_0 \geqq 2$ observe $Y_1, \cdots, Y_{n_0}, Z_1, \cdots, Z_{m_0}$. (b) After $Y_1, \cdots, Y_r$ and $Z_1, \cdots, Z_s$ are observed for $r \geqq n_0$, $s \geqq m_0$, compute $\hat{\alpha}_r, \hat{\beta}_r, \bar{Z}_s, \hat{x}_{(r,s)} = (\bar{Z}_s - \hat{\alpha}_r)/\hat{\beta}_r$, and $\hat{\delta}_{(r,s)} = v_s/u_r$, where $u_r = (1/(r-2)) \sum_{i=1}^{r} (Y_i - \hat{\alpha}_r - \hat{\beta}_r x_i)^2$, $v_s = (1/(s-1)) \sum_{j=1}^{s} (Z_j - \bar{Z}_s)^2$ are the estimators of σ_1^2 and σ_2^2 respectively. (c) For the next observation, observe Y_{r+1} if $(r/s) \leqq ((1 + \hat{x}^2_{(r,s)})/\hat{\delta}_{(r,s)})^{\frac{1}{2}}$; otherwise observe Z_{s+1}. (d) Stop when $r + s = T$ with $N = r$, $M = s$ where N, $M = T - N$ are the random sample sizes. Compute $\hat{x}_N = (\bar{Z}_{T-N} - \hat{\alpha}_N)/\hat{\beta}_N$ and construct a confidence interval $(\hat{x}_N - d_T, \hat{x}_N + d_T)$.

Now letting $\Theta = N/T$ (a random variable) denote the proportion of observations allocated for the Y_i's, we investigate the asymptotic properties of the sequential procedure in the following.

THEOREM 3. *Under the Procedure* R, *if the conditions stated in Theorem* 2 *are satisfied, then*

$$\lim_{T\to\infty}\Theta = \theta_0 \quad \text{a.s.}, \qquad \lim_{T\to\infty} E\Theta = \theta_0, \tag{3.1}$$

$$\lim_{T\to\infty}\{P[|\hat{x}_N - x| \leqq d_T] - g(\theta_0)\} = 0, \tag{3.2}$$

where θ_0 *and* $g(\theta_0)$ *are defined in* (2.7) *and* (2.8), *respectively.*

PROOF. Clearly, as $T\to\infty$, $N\to\infty$ a.s. and $(T-N)\to\infty$ a.s. It follows that $\hat{\alpha}_N\to\alpha$ a.s. and $\hat{\beta}_N\to\beta$ a.s. ([8]), $u_N\to\sigma_1^2$ a.s. and $v_{T-N}\to\sigma_2^2$ a.s. ([4]) as $T\to\infty$; which implies $\hat{\delta}_{(N,T-N)}\to\delta$ a.s. and $\hat{x}_N\to x$ a.s. Applying a lemma in [9] it follows that $\Theta/(1-\Theta) = N/(T-N)\to((1+x^2)/\delta)^{\frac{1}{2}}$ a.s., hence $\Theta\to\theta_0$ a.s., as $T\to\infty$. Since Θ is uniformly bounded, we have $E\Theta\to\theta_0$. This proves (3.1).

It remains to show (3.2). Clearly for every fixed T

$$P[|\hat{x}_N - x| \leqq d_T] = P[|V_T| \leqq T^{\frac{1}{2}}d_T|\hat{\beta}_N|/(\sigma_1\tau_0^{\frac{1}{2}})],$$

where

$$V_T = T^{\frac{1}{2}}(\bar{Z}_{T-N} - \hat{\alpha}_N - x\hat{\beta}_N)/(\sigma_1\tau_0^{\frac{1}{2}}). \tag{3.3}$$

Since $T^{\frac{1}{2}}d_T|\hat{\beta}_N|/(\sigma_1\tau_0^{\frac{1}{2}})$ converges to $a\beta/(\sigma_1\tau_0^{\frac{1}{2}})$ a.s. as $T\to\infty$, by the Slutsky theorem ([3], page 254) it suffices to prove that V_T has an asymptotically standard normal distribution as $T\to\infty$. Let $K = [\theta_0 T]$ denote the largest integer less than or equal to $\theta_0 T$, V_T can be rewritten as

$$\begin{aligned} V_T &= T^{\frac{1}{2}}\{(\bar{Z}_{T-K} - \hat{\alpha}_K - x\hat{\beta}_K) + (\bar{Z}_{T-N} - \bar{Z}_{T-K}) - (\hat{\alpha}_N - \hat{\alpha}_K) \\ &\qquad - x(\hat{\beta}_N - \hat{\beta}_K)\}/(\sigma_1/\tau_0^{\frac{1}{2}}) \\ &= U_{1,T} + U_{2,T} - U_{3,T} - U_{4,T}, \quad \text{say.} \end{aligned} \tag{3.4}$$

By Theorem 2, $U_{1,T}$ is asymptotically normal (0, 1). Therefore again by the Slutsky theorem is suffices to show that $U_{i,T}(i = 2, 3, 4)$ converges to 0 in probability as $T\to\infty$.

We now show the convergence of $U_{4,T}$. Since

$$(\hat{\beta}_n - \beta)/\sigma_1 = \sum_{i=1}^n \frac{(x_i - \bar{x}_n)}{S_n^2} V_i$$

holds for every $n \geqq 3$, where $V_1, V_2, \cdots$ is a sequence of i.i.d. random variables with 0 mean and unit variance, we can write

$$\begin{aligned} T^{\frac{1}{2}}(\hat{\beta}_n - \hat{\beta}_K)/\sigma_1 &= Q\{(R-1)\textstyle\sum_1^I x_i V_i - (R\bar{x}_n - \bar{x}_K)\sum_1^I V_i\} + W_3 \\ &= W_1 + W_2 + W_3, \quad \text{say,} \end{aligned} \tag{3.5}$$

where $I = \min(n, K)$, $Q = T^{\frac{1}{2}}/S_K^2$, $R = S_K^2/S_n^2$ and

$$\begin{aligned} W_3 &= -Q[\textstyle\sum_{n+1}^K x_i V_i - \bar{x}_K \sum_{n+1}^K V_i] && n < K, \\ &= 0 && \text{for} \quad n = K, \\ &= \frac{Q}{R}\textstyle\sum_{K+1}^n x_i V_i - \bar{x}_n \sum_{K+1}^n V_i] && n > K. \end{aligned} \tag{3.6}$$

For arbitrary but fixed $\varepsilon > 0$ let n^* be large enough such that for $n \geqq n^*$,

$$|(S_n^2/n) - 1| < \varepsilon\,, \qquad |(\textstyle\sum_1^n x_i^2/n) - 1| < \varepsilon\,,$$

hold (the existence of n^* is assured by the conditions imposed in Section 2). Let

$$A = A(\varepsilon, T) = \{n \mid n^* \leqq n \leqq T, |(n/K) - 1| < \varepsilon, |R - 1| < \varepsilon^{\frac{3}{2}}\}\,,$$

then by (3.1) there exists a T_0 such that, under the Procedure R,

$$P[N \in A(\varepsilon, T)] \geqq 1 - \varepsilon$$

for every $T > T_0$. Since

$$\begin{aligned} P[\max_{n\in A} |W_1| > \varepsilon] &\leqq P[Q\varepsilon^{\frac{3}{2}} \cdot \max_{n\in A} |\textstyle\sum_1^I x_i V_i| > \varepsilon] \\ &\leqq \varepsilon T \cdot S_H^2/(S_K^2)^2 \leqq \varepsilon(1 + \varepsilon)/(\theta_0(1 - \varepsilon)) = b\varepsilon \quad \text{say,} \end{aligned}$$

where $H = [(1 + \varepsilon)K]$ and the second inequality follows from Kolmogorov's inequality, it follows that

$$\begin{aligned} P[|W_1| > \varepsilon] &\leqq P[\max_{n\in A} |W_1| > \varepsilon, N \in A] + P[N \notin A] \\ &\leqq P[\max_{n\in A} |W_1| > \varepsilon] + \varepsilon < (b + 1)\varepsilon \end{aligned}$$

holds for every $T > T_0$. Therefore $W_1 \rightarrow_p 0$. Similarly it can be shown that $W_2 \rightarrow_p 0$ and $W_3 \rightarrow_p 0$. This implies $U_{4,T} \rightarrow_p 0$ as $T \rightarrow \infty$.

To show the convergence of $U_{2,T}$ consider the expression

$$\begin{aligned} U_{2,T}/\sigma_2 = &\left(1 - \frac{T - N}{T - K}\right) T^{\frac{1}{2}} \textstyle\sum_1^{T-N} V_j/(T - N) \\ &+ T^{\frac{1}{2}}(\textstyle\sum_1^{T-N} V_j - \sum_1^{T-K} V_j)/(T - K)\,, \end{aligned}$$

the assertion follows by the convergence of $(T - N)/(T - K)$ to 1 in probability, the discussion in [2] (page 198) and Kolmogorov's inequality. The convergence of $U_{3,T}$ can be shown similarly. This completes the proof of the theorem.

4. Monte Carlo results and some concluding remarks. The Procedure R had been programmed and Monte Carlo studies on an IBM 360/65 at the University of Nebraska Computing Center were carried out with various sets of parameter values. In most cases the numerical results are quite similar. Table 1 gives the

TABLE 1
Monte Carlo result

	$T = 25$	$T = 50$	$T = 75$	$T = 100$
Normal Distribution				
Average θ value	0.7238	0.7349	0.7305	0.7307
S. D. of θ	0.0888	0.0716	0.0535	0.0412
Observed probability	0.7800	0.8300	0.8350	0.8600
Uniform Distribution				
Average θ value	0.6816	0.6871	0.6779	0.6812
S. D. of θ	0.0998	0.0689	0.0398	0.0241
Observed probability	0.8100	0.8200	0.8000	0.8200

average θ values, their standard deviations and the observed probabilities of coverage of 200 experiments with $\alpha = 0.2$, $\beta = 0.4$, $x = 0.9$, $\sigma_1 = 0.3$, $\sigma_2 = 0.15$, $d_T = 2/T^{\frac{1}{4}}$ and $x_i = (-1)^i$ for $i = 1, 2, \cdots$. Both normal errors and uniform (0, 1) errors were considered in the study.

With this set of parameters $\theta_0 = 0.7291$ and $g(\theta_0) = 0.8516$. It appears that the numerical results and the rates of convergence are acceptable from a practical point of view.

If the situation does not allow the experiment to be carried out sequentially or if the experimenter prefers to apply a single stage procedure, then (2.7) can provide bounds on θ_0 if the experimenter has an idea about the ranges of x and δ; this is because θ_0 is monotonically increasing in $|x|$ and monotonically decreasing in δ. In particular if $\delta = 1(\sigma_1 = \sigma_2)$, then $\theta_0 \geqq \frac{1}{2}$ always holds, and $\theta_0 = \frac{1}{2}$ holds iff $x = 0$. In this case we should always observe more Y_i's than Z_j's, which is not intuitively obvious.

REFERENCES

[1] ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.

[2] CHUNG, K. L. (1968). *A Course in Probability Theory*. Harcourt, Brace and World, New York.

[3] CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.

[4] GLESER, L. J. (1965). On the asymptotic theory of fixed-size sequential confidence bounds for linear regression parameters. *Ann. Math. Statist.* **36** 463-467.

[5] HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests*. Academic Press, New York.

[6] HOADLEY, B. (1970). A Bayesian look at inverse linear regression. *J. Amer. Statist. Assoc.* **65** 356-369.

[7] KRUTCHKOFF, R. G. (1967). Classical and inverse regression methods of calibration. *Technometrics* **9** 425-439.

[8] PERNG, S. K. and TONG, Y. L. (1974). A sequential solution to the inverse linear regression problem. *Ann. Statist.* **2** 535-539.

[9] ROBBINS, H., SIMONS, G. and STARR, N. (1967). A sequential analogue of the Behrens-Fisher problem. *Ann. Math. Statist.* **38** 1384-1391.

[10] SOBEL, M. and TONG, Y. L. (1971). Optimal allocation of observations for partitioning a set of normal populations in comparison with a control. *Biometrika* **58** 177-181.

DEPARTMENT OF STATISTICS
KANSAS STATE UNIVERSITY
MANHATTAN, KANSAS 66502

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF NEBRASKA
LINCOLN, NEBRASKA 68588