NASA TECHNICAL
MEMORANDUM

NASA TM X-58048
June 1970

SMOOTH EMPIRICAL BAYES ESTIMATION WITH APPLICATION TO THE WEIBULL DISTRIBUTION

A Dissertation Presented to the
Graduate Faculty of Texas Tech University
In Partial Fulfillment of
the Requirements for the Degree
Doctor of Philosophy in Industrial Engineering

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
MANNED SPACECRAFT CENTER
HOUSTON, TEXAS

TM X-58048

SMOOTH EMPIRICAL BAYES ESTIMATION WITH APPLICATION TO THE WEIBULL DISTRIBUTION

G. Kemble Bennett
Manned Spacecraft Center
Houston, Texas

ACKNOWLEDGEMENTS

I am deeply indebted to Dr. Harry F. Martz for his direction of this research and to the other members of my committee, Dr. Charles L. Burford, Dr. Brian K. Lambert, Dr. Arun G. Walvekar, and Dr. Thomas L. Boullion for their assistance.

I am also grateful to the National Aeronautics and Space Administration for providing support for this research under NASA Grant NGR 44-011-029. In particular, I wish to thank Dr. J. M. Lewallen of the Theory and Analysis Office for serving as my Academic Supervisor while I was in residence at the Manned Spacecraft Center.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

Figure

Figure

Figure

CHAPTER I

INTRODUCTION AND PRELIMINARIES

In most situations the parameters of a statistical distribution are not known with certainty and must be estimated. In this dissertation these parameters are treated as random variables and empirical Bayes point estimates are given. In particular, the two-parameter Weibull distribution is considered.

1.1 Historical Background of the Weibull Distribution

In 1939 a Swedish scientist, Waloddi Weibull, derived a statistical distribution with which his name has been associated in recent years. This derivation came about as the result of an analysis of breaking-strength data and can be found in [34]. Weibull also published related papers [35], [37], and in [36] illustrates several examples of the distribution's practical value in analyzing various types of data.

The wide audience these papers found among reliability engineers after World War II firmly attached the Weibull name to this statistical distribution. The distribution, however, was originally derived in 1928 by R. A. Fisher and L. H. C. Tippett [8]. Their derivation became known to researchers who were familiar with extreme-value theory as the Fisher-Tippett Type III distribution of extreme values and as the third asymptotic distribution of extreme values.

1.2 Brief Survey of Previous Research

The problem of estimating the parameters in the Weibull distribution has received considerable attention in recent literature from several authors. The techniques proposed by these authors encompass a wide spectrum of statistical methods. These methods will be referred to as "classical" methods of estimation.

Graphical techniques for grouped and ungrouped data have been proposed by Kao [15]. Best linear unbiased estimators (BLUE) were computed by Govinarajulu and Joski [9] using ordered observations for small sample sizes. White [33] obtained linear, unbiased, least squares estimators for the censored Log-Weibull Distribution. Gumbel [10], Menon [21], Miller and Freund [22], and Bain and Antle [1] all give simple estimators, that is, estimators which do not require tedious computations. Maximum-likelihood estimators which generally provide useful estimates have been considered by several authors; among them are Cohen [4], Dubey [7], Harter and Moore [12], and Thoman et al. [31]. Although this by no means represents an exhaustive list of authors concerned with parameter estimation in the Weibull distribution, it does point out the considerable attention that the distribution has received.

1.3 Purpose

The type of decision problem to be considered in this dissertation can best be illustrated by an example. Consider the development program for a particular solid-propellant-rocket engine which must "burn" for a specified time. In this program, certain points exist at which progress is monitored. For instance, the Pre-Flight Rating Test program would be one such point at the culmination of the initial R&D program, demonstrating the ability of a sample of engines to perform for a specified length of time. After this phase a new phase is entered in which flight and static tests are performed, and if needed, a more refined system configuration is developed. Finally, design is frozen, and a Qualification Test program is undertaken to demonstrate the suitability of the engine system. During this period in the program, several groups of engines are test-fired, and due to stringent reliability requirements, a large sample of engines is required.

Throughout these development phases, it is quite possible that the form of the time-to-failure distribution remains unchanged, that is, no significant design changes were incorporated which would greatly affect the overall characteristic performance of the engines. In each of the experiments conducted throughout the total program, however, the classical estimates obtained for the unknown parameters varied unpredictably from experiment to experiment. Since no specific cause for this variation could be found, the variation was considered to be random. This random variation may have been caused by the interaction of the components comprising the engine system, by variation in the solid propellant mixing process, or by numerous other uncontrollable factors.

If the researcher is using a classical estimation procedure during the Qualification Test program, he must choose a large sample size in order to meet the stringent reliability requirement. He would, of course, like to use the data obtained from the previous experiments. His procedure, however, restricts him to the use of only the data in the present experiment. He could consider "pooling" all previous data to obtain point estimates for the parameters; however, he would be violating the basic principles on which classical methods are established and might obtain inaccurate results.

To use these methods, it is necessary to assume that the data are obtained from the same specified distribution. Therefore, in order to pool data from previous experiments, the unknown parameters must remain constant throughout each experiment. This is not usually the case in a development program; therefore previous data must be ignored when using classical techniques.

Since the fluctuation in the parameters for the time-to-failure distribution can be attributed to random variation, data from previous experiments can and should be used to obtain point estimates in the present experiment. When the researcher knows the *prior* distribution describing this variation, the Bayes principle can be used to provide estimates for the parameters. In terms of the minimization of the overall expectation of some appropriate loss structure, this principle provides "best" estimates for these parameters. Two attempts to apply Bayes analysis when the time-to-failure distribution is Weibull are given by Soland [30] and Harris and Singpurivalla [11]. In each of these papers various forms for the prior distribution are considered.

The Bayes method for parameter estimation is difficult to apply in many situations. For "best" results it requires a completely known and specified prior distribution, which is seldom available. In such situations an empirical Bayes approach can be utilized. This approach acknowledges the existence of a prior distribution; however, this distribution need never be explicitly known to the researcher. Estimates of the parameters obtained in previous experiments can be used to improve the estimates in the present experiment. This approach will be discussed in detail in the next section.

The main purpose of this dissertation is to develop an empirical Bayes estimator that is capable of providing significant improvement over both the classical and presently known empirical Bayes estimators. The estimator will then be used to obtain point estimates for the unknown parameters in the two-parameter Weibull distribution. The particular form of the distribution considered has the following probability density function:

$$f(x) \;=\; \alpha\beta x^{\beta-1}\, e^{-\alpha x^{\beta}} \qquad (x \geq 0;\ \alpha,\ \beta > 0) \qquad (1.1)$$

where α denotes the *scale* parameter and β the *shape* parameter. Figures 1 and 2 illustrate the influence these parameters have on the Weibull density function (1.1). In Figure 1, $\alpha = 1$ and plots of f(x) in (1.1) are given for various values of β . We remark that for $\beta = 1$, f(x) plots as a one-parameter exponential density function with the parameter equal to one. Thus, when $\beta = 1$, (1.1) reduces to the well-known exponential density function with parameter α . In Figure 2, $\beta = 3$ and plots of f(x) from (1.1) are given for several values of α .

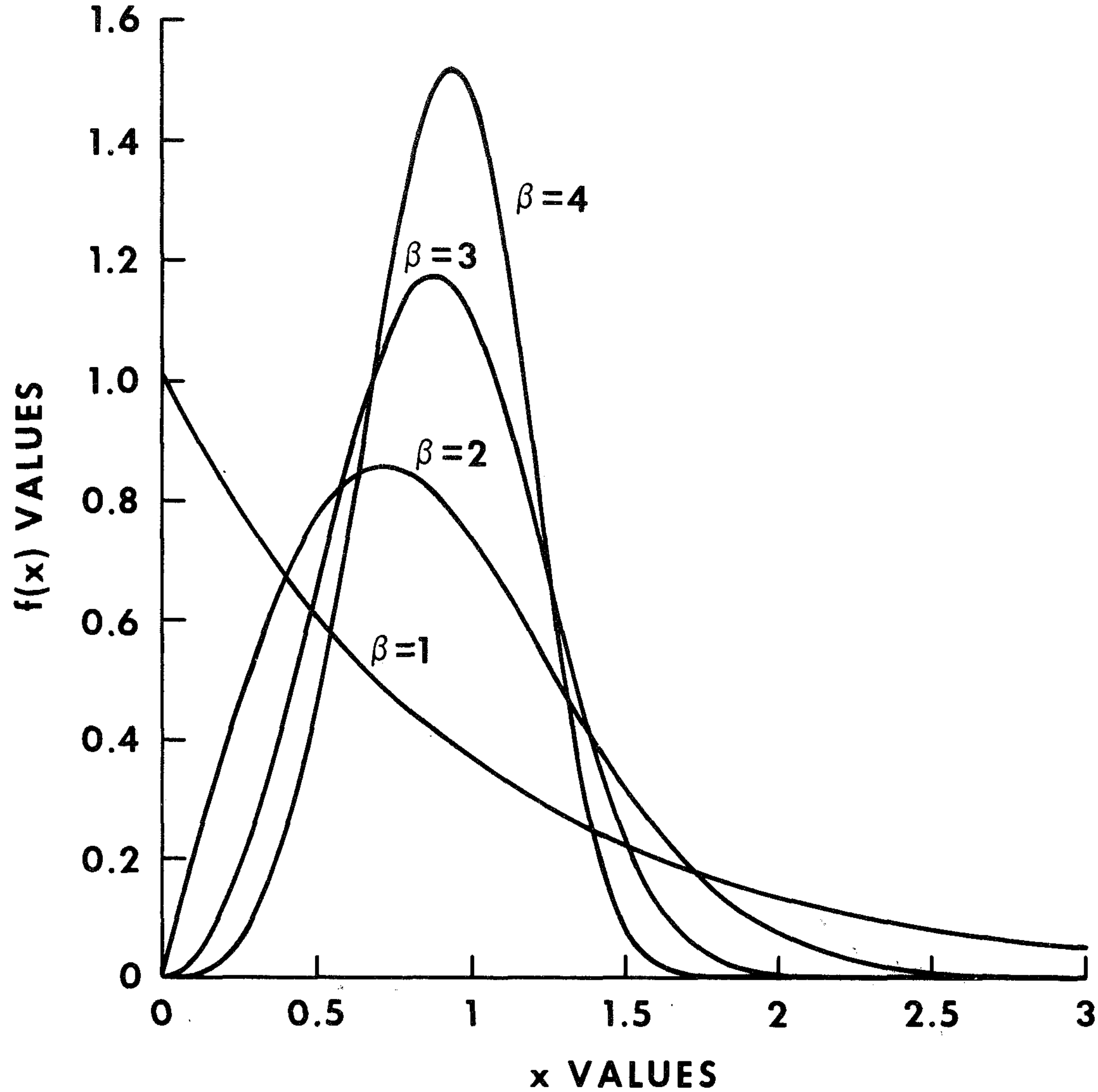

Figure 1. — Weibull density function (α = 1).

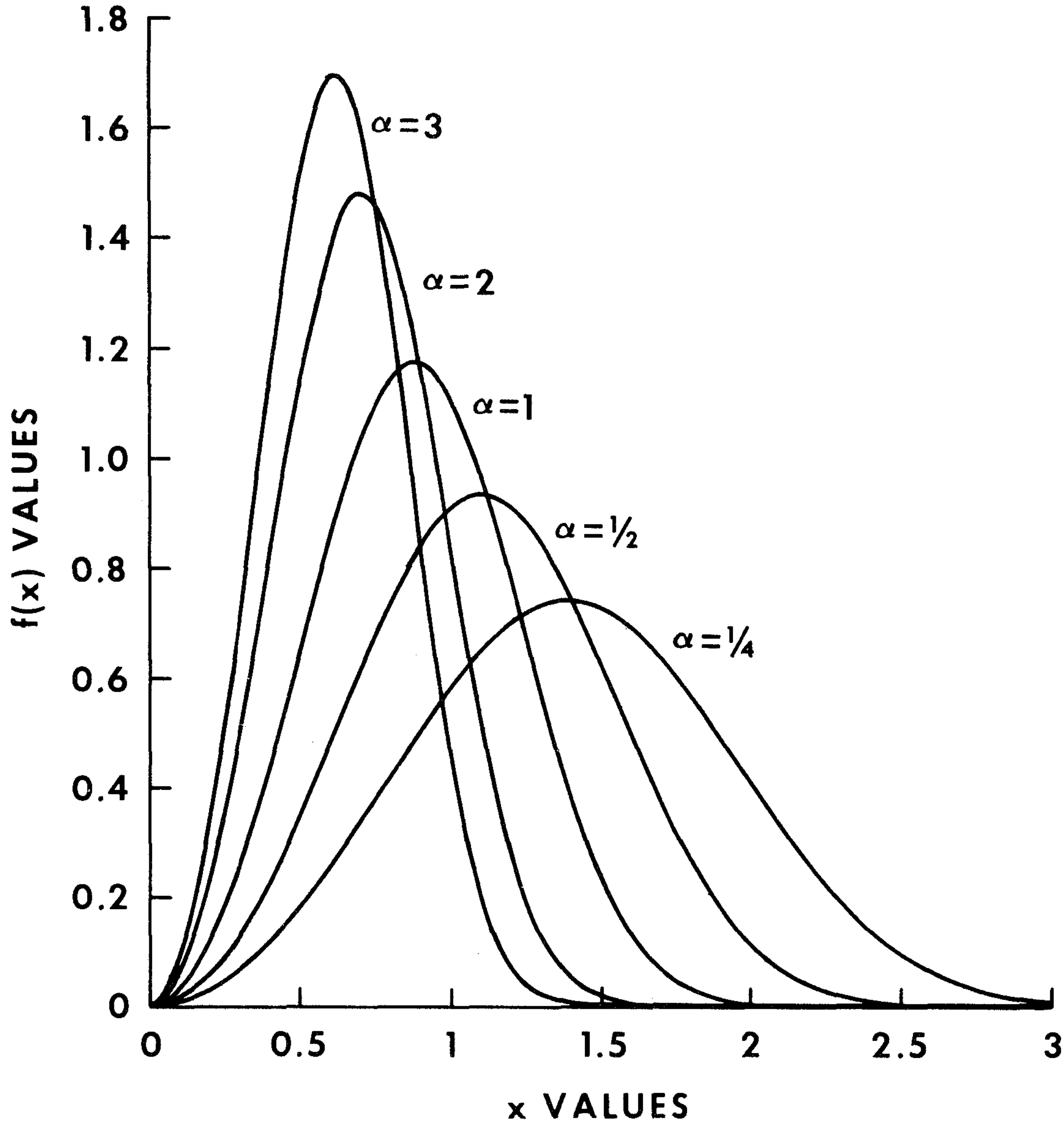

Figure 2. — Weibull density function (β = 3).

1.4 Bayes and Empirical Bayes Estimation

Throughout this dissertation, upper-case Greek or Roman letters will identify random variables. Lower-case letters will be reserved for realizations of the same. In some cases ease of notation may cause violation of this convention; however, in such instances proper meaning will be clear.

To begin let us define the basic-decision theoretic elements on which both the empirical Bayes and the Bayes approach are based. They are as follows:

(i) There is a parameter space $\underset{\sim}{\Theta}$ with generic s-vector $\underset{\sim}{\theta} = (\theta_1, \theta_2, \cdots, \theta_s)$ on which is defined a probability distribution G referred to as a prior distribution.

(ii) There is a decision space D, which coincides with $\underset{\sim}{\Theta}$ for estimation, with generic element $\underset{\sim}{\delta}$.

(iii) There is a loss function $\ell(\underset{\sim}{\delta}, \underset{\sim}{\theta}) \geq 0$ representing the loss incurred when $\underset{\sim}{\delta}$ is taken as an estimate for $\underset{\sim}{\theta}$.

(iv) There is an observable random k-vector $\underset{\sim}{x} = (x_1, x_2, \cdots, x_k)$, distributed on a space $\underset{\sim}{X}$ on which is defined a σ-finite measure μ.

When the parameter is $\underset{\sim}{\theta}$, $\underset{\sim}{X}$ has a density $f(\underset{\sim}{x}|\underset{\sim}{\theta})$ with respect to μ .

If the decision function $\underset{\sim}{\delta}$ is chosen and the sample vector is observed, then $\underset{\sim}{\delta}(\underset{\sim}{x})$ is taken as an estimate of $\underset{\sim}{\theta}$ and the loss $\ell(\underset{\sim}{\delta}(\underset{\sim}{x}), \underset{\sim}{\theta})$ is incurred. For any such decision $\underset{\sim}{\delta}$ the expected loss when $\underset{\sim}{\theta}$ is the true parameter is given by

$$R(\underset{\sim}{\delta}, \underset{\sim}{\theta}) = \int_{\underset{\sim}{X}} \ell(\underset{\sim}{\delta}(\underset{\sim}{x}), \underset{\sim}{\theta}) f(\underset{\sim}{x}|\underset{\sim}{\theta})\, d\mu(\underset{\sim}{x}) \quad . \tag{1.2}$$

Hence, the global or overall risk can be represented as

$$R(\underset{\sim}{\delta}, G) = \int_{\underset{\sim}{\Theta}} R(\underset{\sim}{\delta}, \underset{\sim}{\theta})\, dG(\underset{\sim}{\theta}) \tag{1.3}$$

where G is the prior distribution of $\underset{\sim}{\Theta}$. If the prior density corresponding to $G(\underset{\sim}{\theta})$ is denoted by $g(\underset{\sim}{\theta})$, this risk can be expressed as

$$R(\underset{\sim}{\delta}, G) = \int_{\underset{\sim}{X}} f(\underset{\sim}{x}) \int_{\underset{\sim}{\Theta}} \ell(\underset{\sim}{\delta}(\underset{\sim}{x}), \underset{\sim}{\theta}) h(\underset{\sim}{\theta}|\underset{\sim}{x})\, d\underset{\sim}{\theta}\, d\underset{\sim}{x}$$

where

$$h(\underset{\sim}{\theta}|\underset{\sim}{x}) = \frac{f(\underset{\sim}{x}|\underset{\sim}{\theta})\; g(\underset{\sim}{\theta})}{f(\underset{\sim}{x})} = \frac{f(\underset{\sim}{x}, \underset{\sim}{\theta})}{f(\underset{\sim}{x})} \tag{1.4}$$

and $f(\underset{\sim}{x})$ is the marginal density of $\underset{\sim}{X}$. Therefore, to minimize the risk $R(\underset{\sim}{\delta}, G)$, a decision function $\underset{\sim}{\delta}(\underset{\sim}{x})$ should be chosen such that

$$\phi(\underset{\sim}{\delta}, \underset{\sim}{x}) = \int_{\underset{\sim}{\Theta}} \ell\left(\underset{\sim}{\delta}(\underset{\sim}{x}), \underset{\sim}{\theta}\right) h(\underset{\sim}{\theta}|\underset{\sim}{x})\, d\underset{\sim}{\theta} \qquad (1.5)$$

is minimum. If such a decision function, say $\underset{\sim}{\delta}_G(\underset{\sim}{x})$, exists it is known as the Bayes decision function and

$$R(G) = R(\underset{\sim}{\delta}_G, G) = \min_{\underset{\sim}{\delta}}\left(R(\underset{\sim}{\delta}, G)\right) \qquad (1.6)$$

is known as the Bayes risk.

When G is known, the optimal decision $\underset{\sim}{\delta}_G$ can be determined. If G is unknown, however, this minimum risk decision cannot be obtained. In the empirical Bayes approach complete determination and specification of the prior distribution is unnecessary. Instead, it is assumed that the decision problem given above has occurred repeatedly and independently with the same unknown prior distribution throughout. Thus, there exists a sequence

$$(\underset{\sim}{X}_1, \underset{\sim}{\theta}_1), (\underset{\sim}{X}_2, \underset{\sim}{\theta}_2), \cdots, (\underset{\sim}{X}_n, \underset{\sim}{\theta}_n) \qquad (1.7)$$

of independent pairs of random vectors $(\underset{\sim}{X}, \underset{\sim}{\theta})$, where $\underset{\sim}{X}_i (i = 1,2,\cdots,n)$ has dimension k and $\underset{\sim}{\theta}_i (i = 1,2,\cdots,n)$

has dimension s . At the time an estimate of $\underset{\sim}{\Theta}$ for the nth realization or present experiment is to be determined, the researcher has at his disposal the vector-valued sequence $\underset{\sim}{x}_1, \underset{\sim}{x}_2, \cdots, \underset{\sim}{x}_n$. The s-vector $\underset{\sim}{\theta}$ remains, of course, unknown. This information can be used to provide a decision function of $\underset{\sim}{x}_n$ based upon $\underset{\sim}{x}_1, \underset{\sim}{x}_2, \cdots, \underset{\sim}{x}_{n-1}$ like

$$\underset{\sim}{\delta}_n(\underset{\sim}{x}_n) = \underset{\sim}{\delta}_n(\underset{\sim}{x}_1, \underset{\sim}{x}_2, \cdots, \underset{\sim}{x}_{n-1}; \underset{\sim}{x}_n) \tag{1.8}$$

such that when $\underset{\sim}{x}_n$ is observed, $\underset{\sim}{\delta}_n \varepsilon D$ is taken as an estimate of $\underset{\sim}{\theta}_n$ and the loss $\ell\left(\underset{\sim}{\delta}_n(\underset{\sim}{x}_n), \underset{\sim}{\theta}_n\right)$ incurred. Such a decision function will be called an empirical Bayes decision function. Hence, when G is unknown to the researcher, he is able to extract some information about the prior distribution through the sequence $\underset{\sim}{x}_1, \underset{\sim}{x}_2, \cdots, \underset{\sim}{x}_n$ and obtain an approximate decision function $\underset{\sim}{\delta}_n(\underset{\sim}{x}_n)$ to the Bayes decision function $\underset{\sim}{\delta}_G(\underset{\sim}{x})$.

1.5 Squared Error Loss

Consider the squared error loss function

$$\ell\left(\delta_i(\underset{\sim}{x}), \theta_i\right) = \left(\delta_i(\underset{\sim}{x})-\theta_i\right)^2 \tag{1.9}$$

for the θ_i component of $\underset{\sim}{\Theta}$. If we represent $\phi_G(\delta_i, \underset{\sim}{x})$ defined by (1.5) as $E\left[\ell\left(\delta_i(\underset{\sim}{x}), \theta_i\right)|\underset{\sim}{x}\right]$ and replace

$\ell\left(\delta_i(\underset{\sim}{x}),\ \theta_i\right)$ by (1.9) then (1.5) can be written as

$$E\left[\left(\delta_i(\underset{\sim}{x}) - \theta_i\right)^2 \middle| \underset{\sim}{x}\right] \ . \tag{1.10}$$

After adding and subtracting appropriate quantities and simplifying, we have

$$E\left[\left(\delta_i(\underset{\sim}{x}) - \theta_i\right)^2 \middle| \underset{\sim}{x}\right] = \left(\delta_i(\underset{\sim}{x}) - E(\theta_i|\underset{\sim}{x})\right)^2 + \mathrm{Var}(\theta_i|\underset{\sim}{x}) \tag{1.11}$$

which will be the minimum when

$$\delta_i(\underset{\sim}{x}) = E(\theta_i|\underset{\sim}{x}) \ . \tag{1.12}$$

The Bayes estimator for each $\Theta_i (i = 1, 2, \cdots, s)$ is therefore given by (1.12), and from (1.6) the Bayes risk becomes $E\left[\mathrm{Var}(\theta_i|\underset{\sim}{x})\right]$. In general, the Bayes estimator for $\underset{\sim}{\Theta}$ is given by

$$\underset{\sim}{\delta}(\underset{\sim}{x}) = E(\underset{\sim}{\theta}|\underset{\sim}{x}) \ . \tag{1.13}$$

1.6 Bayes Estimate for a Sufficient Statistic

Consider an arbitrary random sample of size k denoted by $\underset{\sim}{x} = (x_1, x_2, \cdots, x_k)$ from a univariate distribution with density function $f(x|\theta)$. If there exists a set of sufficient statistics $\underset{\sim}{t} = (t_1, t_2, \cdots, t_s)$ for $\underset{\sim}{\theta} = (\theta_1, \theta_2, \cdots, \theta_s)$, then by the Neyman factorization criterion [23] the *likelihood function*

$$L(\underset{\sim}{x}|\underset{\sim}{\theta}) = \prod_{i=1}^{k} f(x_i|\underset{\sim}{\theta}) \tag{1.14}$$

factors into

$$L(\underset{\sim}{x}|\underset{\sim}{\theta}) = f(\underset{\sim}{t}|\underset{\sim}{\theta})\ c(\underset{\sim}{x}) \tag{1.15}$$

where $f(\underset{\sim}{t}|\underset{\sim}{\theta})$ represents the conditional density function of $\underset{\sim}{t}$ and $c(\underset{\sim}{x})$ is some function of $\underset{\sim}{x}$ not involving $\underset{\sim}{\theta}$. If the prior density function is denoted by $g(\underset{\sim}{\theta})$, the joint density function of $\underset{\sim}{x}$ and $\underset{\sim}{\theta}$ can be written

$$f(\underset{\sim}{x}, \underset{\sim}{\theta}) = L(\underset{\sim}{x}|\underset{\sim}{\theta})\ g(\underset{\sim}{\theta}) ,$$

and by (1.15) this becomes

$$\begin{aligned} f(\underset{\sim}{x}, \underset{\sim}{\theta}) &= f(\underset{\sim}{t}|\underset{\sim}{\theta})\ g(\underset{\sim}{\theta})\ c(\underset{\sim}{x}) \\ &= f(\underset{\sim}{t}, \underset{\sim}{\theta})\ c(\underset{\sim}{x}) . \end{aligned} \tag{1.16}$$

If (1.16) is integrated over the region $\underset{\sim}{\Theta}$, the marginal density of $\underset{\sim}{X}$ becomes

$$f(\underset{\sim}{x}) = p(\underset{\sim}{t})c(\underset{\sim}{x}) \tag{1.17}$$

and by dividing (1.17) into (1.16), we obtain

$$h(\underset{\sim}{\theta}|\underset{\sim}{x}) = \frac{f(\underset{\sim}{t},\underset{\sim}{\theta})}{p(\underset{\sim}{t})} = f(\underset{\sim}{\theta}|\underset{\sim}{t}) .$$

Hence, the Bayes estimator for $\underset{\sim}{\theta}$ becomes

$$E(\underset{\sim}{\theta}|\underset{\sim}{x}) = E(\underset{\sim}{\theta}|\underset{\sim}{t}) \tag{1.18}$$

and the empirical Bayes estimator can be based on $\underset{\sim}{\tau} = (\underset{\sim}{t}_1, \underset{\sim}{t}_2, \cdots, \underset{\sim}{t}_n)$ rather than $\underset{\sim}{X} = (\underset{\sim}{x}_1, \underset{\sim}{x}_2, \cdots, \underset{\sim}{x}_n)$. Thus, the sequence (1.7) which represents past experience can be written as

$$(\underset{\sim}{T}_1, \underset{\sim}{\theta}_1), (\underset{\sim}{T}_2, \underset{\sim}{\theta}_2) \cdots (\underset{\sim}{T}_n, \underset{\sim}{\theta}_n) \tag{1.19}$$

with each vector pair $(\underset{\sim}{T}, \underset{\sim}{\theta})$ distributed identically and independently with probability density $f(\underset{\sim}{t}|\underset{\sim}{\theta})g(\underset{\sim}{\theta})$.

1.7 Maximum-Likelihood Estimation

In this section several properties of maximum-likelihood estimation are presented for completeness. These properties will be needed in the remaining chapters and can essentially be found in Kendall and Stuart [16].

Throughout this section the reader is reminded that the parameter $\underset{\sim}{\theta}$, specifying the particular member of the family of distributions under consideration, is not assumed to be a random variable. Thus, notation such as $f(x|\underset{\sim}{\theta})$ merely exemplifies the dependence of the density on the parameter $\underset{\sim}{\theta}$ and should not be confused with the analagous conditional probability statement. This notation is followed, since in the remaining sections, $\underset{\sim}{\theta}$ is assumed to be a random variable.

The *maximum-likelihood estimate* of $\underset{\sim}{\theta}$ is defined as that value, say $\hat{\underset{\sim}{\theta}}_k$, within the range of $\underset{\sim}{\Theta}$ which maximizes the likelihood function (1.14). The subscript k is used to denote the dependence of the estimate on the sample size. If the likelihood function L is a twice differentiable function of $\underset{\sim}{\Theta}$ throughout its range and if stationary values of $L(\underset{\sim}{x}|\underset{\sim}{\theta})$ exist, then the maximum-likelihood estimates of $\theta_1, \theta_2, \cdots, \theta_s$ can be found by solution of the system of equations

$$\frac{\partial L(\underset{\sim}{x}|\underset{\sim}{\theta})}{\partial \theta_i} = 0 , \qquad (i = 1,2,\cdots,s) , \qquad (1.20)$$

for $\hat{\theta}_{k,1}, \hat{\theta}_{k,2}, \cdots, \hat{\theta}_{k,s}$.

In practice it is often simpler to work with the logarithm of the likelihood function rather than the function itself. Since $L(\underset{\sim}{x}|\underset{\sim}{\theta})$ and $\log L(\underset{\sim}{x}|\underset{\sim}{\theta})$ have their maximum at the same value of $\underset{\sim}{\theta}\varepsilon\underset{\sim}{\Theta}$. Thus, maximum-likelihood estimates can be found by the solution of the system of equations

$$\frac{\partial \log L(\underset{\sim}{x}|\underset{\sim}{\theta})}{\partial \theta_i} = 0 , \qquad (i = 1,2,\cdots,s) . \qquad (1.21)$$

When $s = 1$, (1.21) reduces simply to

$$\frac{d \log L(\underset{\sim}{x}|\theta)}{d\theta} = 0 . \qquad (1.22)$$

Consider the univariate density function $f(x|\underset{\sim}{\theta})$. If a sufficient statistic $\underset{\sim}{t}$ exists for the parameter vector $\underset{\sim}{\theta}$ it is readily seen that the maximum-likelihood estimator $\hat{\underset{\sim}{\theta}}_k$, must be a function of it. This is true since the sufficiency of $\underset{\sim}{t}$ for $\underset{\sim}{\theta}$ implies the factorization given by (1.15). Choosing $\hat{\underset{\sim}{\theta}}_k$ to maximize the likelihood function is thus equivalent to choosing $\hat{\underset{\sim}{\theta}}_k$ to maximize $f(\underset{\sim}{t}|\underset{\sim}{\theta})$, and hence $\hat{\underset{\sim}{\theta}}_k$ is a function of $\underset{\sim}{t}$ alone. If $\hat{\underset{\sim}{\theta}}_k$ represents a one-to-one transformation on $\underset{\sim}{t}$, then $\hat{\underset{\sim}{\theta}}_k$ will be sufficient for $\underset{\sim}{\theta}$.

In a proof given by Wald [32], the MLE is shown to be consistent under quite general conditions. A simplified version of this proof is available in [16]. The generality of the conditions becomes clear by the absence of any regularity constraints on the density function $f(x|\underset{\sim}{\theta})$. These conditions require the existence of certain integrals which are generally satisfied by most distributions and, in particular, by the various forms of the Weibull distribution.

Consider the case where $s = 1$. If the first two derivatives of the likelihood function with respect to θ exist, if

$$E\left(\frac{\partial \log L(\underset{\sim}{x}|\theta)}{\partial \theta}\right) = 0 \tag{1.23}$$

and if

$$R^2(\theta) = -E\left(\frac{\partial^2 \log L(\underset{\sim}{x}|\theta)}{\partial \theta^2}\right) = E\left[\left(\frac{\partial \log L(\underset{\sim}{x}|\theta)}{\partial \theta}\right)^2\right] \tag{1.24}$$

exists and is nonvanishing for all $\theta \varepsilon \Theta$, then the maximum-likelihood estimator $\hat{\Theta}_k$ can be shown [16] to be asymptotically normally distributed with mean θ and variance $1/R^2(\theta)$, that is,

$$\text{distr.}\ (\hat{\Theta}_k|\theta) = N\left(\theta, \frac{1}{R^2(\theta)}\right) . \tag{1.25}$$

It will be constructive to show under what conditions the assumptions (1.23) and (1.24) are valid. In this regard consider

$$E\left(\frac{\partial \log L(\underset{\sim}{x}|\theta)}{\partial\theta}\right) = \int_{\underset{\sim}{X}}\left(\frac{1}{L(\underset{\sim}{x}|\theta)}\,\frac{\partial L(\underset{\sim}{x}|\theta)}{\partial\theta}\right)L(\underset{\sim}{x}|\theta)\,d\underset{\sim}{x} \tag{1.26}$$

$$= \int_{\underset{\sim}{X}} \frac{\partial}{\partial\theta} L(\underset{\sim}{x}|\theta)\,d\underset{\sim}{x} \quad .$$

If differentiation and integration can be interchanged, then (1.26) becomes

$$E\left(\frac{\partial \log L(\underset{\sim}{x}|\theta)}{\partial\theta}\right) = \frac{\partial}{\partial\theta}\int_{\underset{\sim}{X}} L(\underset{\sim}{x}|\theta)\,d\underset{\sim}{x} = 0 \quad ,$$

since

$$\int_{\underset{\sim}{X}} L(\underset{\sim}{x}|\theta)\,d\underset{\sim}{x} = 1 \quad . \tag{1.27}$$

Differentiating (1.26) again, we obtain

$$\int_{\underset{\sim}{X}}\left[\left(\frac{1}{L(\underset{\sim}{x}|\theta)}\,\frac{\partial L(\underset{\sim}{x}|\theta)}{\partial\theta}\right)\frac{\partial L(\underset{\sim}{x}|\theta)}{\partial\theta} + L(\underset{\sim}{x}|\theta)\,\frac{\partial}{\partial\theta}\left(\frac{1}{L(\underset{\sim}{x}|\theta)}\,\frac{\partial}{\partial\theta}L(\underset{\sim}{x}|\theta)\right)\right]d\underset{\sim}{x} = 0$$

which becomes

$$\int_{\underset{\sim}{X}} \left[\left(\frac{1}{L(\underset{\sim}{x}|\theta)} \frac{\partial L(\underset{\sim}{x}|\theta)}{\partial \theta} \right)^2 + \frac{\partial^2 \log L(\underset{\sim}{x}|\theta)}{\partial \theta^2} \right] L(\underset{\sim}{x}|\theta)\, d\underset{\sim}{x}$$

or

$$E\left[\left(\frac{\partial \log L(\underset{\sim}{x}|\theta)}{\partial \theta} \right)^2 \right] = -E\left(\frac{\partial^2 \log L(\underset{\sim}{x}|\theta)}{\partial \theta^2} \right) . \qquad (1.28)$$

Thus, the only condition necessary to establish assumptions (1.23) and (1.24) when the first two derivatives of the likelihood function exist is the ability to interchange integration and differentiation. See, for example, Cramér [5].

For the general case $s > 1$, analogous arguments to those above (see [16]) verify that $\hat{\underset{\sim}{\theta}}_k$ is asymptotically distributed as a multivariate normal distribution with mean vector $\underset{\sim}{\theta}$ and covariance matrix V whose inverse V^{-1} has elements given by

$$v_{ij}^{-1} = -E\left[\frac{\partial^2 \log L(\underset{\sim}{x}|\underset{\sim}{\theta})}{\partial \theta_i \partial \theta_j} \right] . \qquad (1.29)$$

1.8 Outline of Succeeding Chapters

In Chapter II a continuously smooth empirical Bayes estimator is developed. This estimator is obtained by replacing the continuous prior density in the Bayes estimator by a suitable approximation. The approximation is based on a sequence of consistent estimates and is shown to converge in probability to the prior density as both the number of past experiments n and the sample size k tend to infinity. Since both n and k are finite for practical application, the mean and variance of the marginal distribution of these estimates may not coincide with those of the prior distribution. Therefore, a method for transforming these estimates into a new sequence of values having a marginal distribution whose mean and variance approximate those of the prior distribution is illustrated. This new sequence of values is then used to obtain an alternative approximation to the prior density.

In Chapter III smooth empirical Bayes estimates are obtained for the discrete Poisson distribution. This chapter has been included in order to exemplify the versatility of the smooth estimation procedure. The distribution has received considerable attention from empirical Bayes authors and can be used to provide a common mode for comparison with other well known and proven empirical Bayes estimators. Two such methods are used for comparison in Chapter VII.

In Chapters IV, V, and VI smooth empirical Bayes estimation is applied to the Weibull distribution. Chapter IV provides smooth empirical Bayes estimators for the scale parameter α when the shape parameter β is known. Chapter V considers the reverse situation, and Chapter VI provides smooth estimators when both parameters are known to vary unpredictably. In each of these chapters, results from Monte Carlo simulation show that even for small sample sizes and few past experiments, the smooth estimators have smaller, mean-squared errors than the classical, maximum-likelihood estimators.

In Chapter VII empirical Bayes methods for point estimation developed by Rutherford and Krutchkoff [28] and Lemon and Krutchkoff [17] are outlined. Where applicable, these methods are applied to the distributions of the preceding chapters, and Monte Carlo simulations are performed. The results from these simulations are then directly compared with the results obtained by using the continuously smooth empirical Bayes estimators.

A summary of the conclusions derived from this research is presented in Chapter VIII, and areas recommended for future research are also discussed.

CHAPTER II

A CONTINUOUSLY SMOOTH EMPIRICAL BAYES ESTIMATOR

In this chapter a continuously smooth empirical Bayes estimator is developed. The estimator is obtained by a continuous approximation to the prior density function and offers a new approach for obtaining empirical Bayes point estimates.

2.1 Notation and Preliminaries

Throughout this dissertation the term consistent estimator will frequently be employed. To say that $\hat{\underset{\sim}{\theta}}_k$ is a *consistent estimator* for $\underset{\sim}{\theta}$ will imply that for any positive numbers ε and δ, however small, there exists an integer K such that for $k > K$

$$\Pr(|\hat{\underset{\sim}{\theta}}_k - \underset{\sim}{\theta}| < \varepsilon) \geq 1 - \delta \quad .$$

This will often be referred to as *convergence in probability* and for notational convenience will be written as

$$\underset{k\to\infty}{\text{p lim}}\ \hat{\underset{\sim}{\theta}}_k = \underset{\sim}{\theta} \quad .$$

Assume that an unobservable random parameter $\underset{\sim}{\theta} = (\theta_1, \theta_2, \cdots, \theta_s)$ occurs according to the distribution $G(\underset{\sim}{\theta})$ with corresponding density function $g(\underset{\sim}{\theta})$. When $\underset{\sim}{\theta}$

is realized, an observable random k-vector $\underset{\sim}{X}$ occurs according to the conditional distribution $F(\underset{\sim}{x}|\underset{\sim}{\theta})$. When a sufficient estimate $\hat{\underset{\sim}{\theta}}_k$ of $\underset{\sim}{\theta}$ exists, the Bayes estimator $E(\theta_j|\underset{\sim}{x})$ for θ_j, the *jth* component of $\underset{\sim}{\theta}$, is given by

$$E(\theta_j|\hat{\underset{\sim}{\theta}}_k) = \frac{\int_{\underset{\sim}{\Theta}} \theta_j \ f(\hat{\underset{\sim}{\theta}}_k|\underset{\sim}{\theta})g(\underset{\sim}{\theta})\ d\underset{\sim}{\theta}}{\int_{\underset{\sim}{\Theta}} f(\hat{\underset{\sim}{\theta}}_k|\underset{\sim}{\theta})g(\underset{\sim}{\theta})\ d\underset{\sim}{\theta}} \tag{2.1}$$

where $f(\hat{\underset{\sim}{\theta}}_k|\underset{\sim}{\theta})$ denotes the conditional density function of the sufficient estimator $\hat{\underset{\sim}{\theta}}_k$ given $\underset{\sim}{\theta}$. Any such function $f(\cdot|\cdot)$ will be referred to as the *kernel of integration.*

When the prior density $g(\underset{\sim}{\theta})$ is unknown, the Bayes estimator is unattainable. However, if the situation described above occurs repeatedly with the same, but unknown $g(\underset{\sim}{\theta})$, then a sequence of n sufficient estimates

$$\hat{\underset{\sim}{\theta}}_{k,1},\ \hat{\underset{\sim}{\theta}}_{k,2},\ \cdots,\ \hat{\underset{\sim}{\theta}}_{k,n} \tag{2.2}$$

obtained from previous replications can be used to construct a continuous approximation to the prior density function. Based on this approximation an empirical Bayes estimate of $\underset{\sim}{\theta}_n$, the *nth* realization from $g(\underset{\sim}{\theta})$, can be determined.

2.2 Marginal Density Approximation

The marginal density function of $\hat{\underset{\sim}{\theta}}_k$ is given by

$$p(\hat{\underset{\sim}{\theta}}_k) = \int_{\underset{\sim}{\Theta}} f(\hat{\underset{\sim}{\theta}}_k | \underset{\sim}{\theta}) g(\underset{\sim}{\theta}) \, d\underset{\sim}{\theta} \quad . \tag{2.3}$$

When $g(\underset{\sim}{\theta})$ is unknown, $p(\hat{\underset{\sim}{\theta}}_k)$ cannot be determined; however, consistent density estimators are available which can be used to approximate $p(\hat{\underset{\sim}{\theta}}_k)$. These estimators can be constructed using sequence (2.2). They take the form

$$p_n(\hat{\underset{\sim}{\theta}}_k) = \frac{1}{nh^s} \sum_{i=1}^{n} W\left(\frac{\hat{\underset{\sim}{\theta}}_k - \hat{\underset{\sim}{\theta}}_{k,i}}{h}\right) \tag{2.4}$$

where h is a function of n satisfying

$$\lim_{n\to\infty} h(n) = 0 \tag{2.5}$$

and

$$\lim_{n\to\infty} nh^s(n) = \infty \quad . \tag{2.6}$$

Acceptable forms for the function W are given by Parzen [24] for univariate densities, $s = 1$, and by Martz [19] for multivariate densities. In particular for estimating a univariate density function, the estimator

$$p_n(\hat{\theta}_k) = \frac{1}{2\pi nh} \sum_{i=1}^{n} \left[\frac{\sin\left(\frac{\hat{\theta}_k - \hat{\theta}_{k,i}}{2h}\right)}{\left(\frac{\hat{\theta}_k - \hat{\theta}_{k,i}}{2h}\right)} \right]^2 \tag{2.7}$$

where

$$h = n^{-1/5} \tag{2.8}$$

will be used. This particular estimator has been demonstrated to possess certain desirable properties (see Clemmer and Krutchkoff [3] and Martz and Krutchkoff [20].

In practice, the estimates $\hat{\theta}_{\sim k,i}$ $(i = 1,2,\cdots,n)$ are unitized quantities, and it becomes necessary to multiply h by an appropriate function to remove these units of measurement from the argument of W in (2.4). For example in (2.7), this may be accomplished by defining h to be

$$h = n^{-1/5} \sqrt{\frac{\sum_{i=1}^{n} (\hat{\theta}_{k,i} - \bar{\theta}_k)^2}{n}} \tag{2.9}$$

where

$$\bar{\theta}_k = \sum_{i=1}^{n} \frac{\hat{\theta}_{k,i}}{n} \quad . \tag{2.10}$$

The estimators $p_n(\hat{\underset{\sim}{\theta}}_k)$ are squared error consistent for estimating the density $p(\hat{\underset{\sim}{\theta}}_k)$ in the sense that

$$\lim_{n\to\infty} E\left[p_n(\hat{\underset{\sim}{\theta}}_k) - p(\hat{\underset{\sim}{\theta}}_k)\right]^2 = 0 \tag{2.11}$$

for all $\hat{\underset{\sim}{\theta}}_k$ in the continuity set of $p(\cdot)$. This convergence implies a different interpretation of squared error consistency. Usually, it is the sample size k which tends to infinity and not the number of experiences n . While increasing sample size is conceptual, in the empirical Bayes situation the number of experiences may actually grow without bound. Thus in application, the convergence given by (2.11) represents a natural result.

2.3 Prior Density Approximation

In this section the marginal density estimator $p_n(\hat{\underset{\sim}{\theta}}_k)$ will be shown to converge in probability to the prior density function $g(\underset{\sim}{\theta})$, as both the number of experiences n and the sample size k tend toward infinity. The following theorem, which is a slightly modified version of a theorem found in Rao [25], will be needed to establish the main result.

Theorem 2.1 – If $p_n(\hat{\underset{\sim}{\theta}}_k)$ is a continuous density estimator and $\hat{\underset{\sim}{\theta}}_k$ is a consistent estimator for $\underset{\sim}{\theta}$, then

$$\underset{k\to\infty}{\text{p lim}}\ p_n(\hat{\underset{\sim}{\theta}}_k) = p_n(\underset{\sim}{\theta}) \quad . \tag{2.12}$$

Proof – Since $\hat{\underset{\sim}{\theta}}_k$ is consistent for $\underset{\sim}{\theta}$, given any positive numbers γ and η, an integer K exists such that for $k > K$ $\Pr(|\hat{\underset{\sim}{\theta}}_k - \underset{\sim}{\theta}| < \gamma) \geq 1 - \eta/2$. Now let I be a finite region such that $\Pr(\underset{\sim}{\theta} \in I) = 1 - \eta/2$. Then since p_n is continuous, $|p_n(\hat{\underset{\sim}{\theta}}_k) - p_n(\underset{\sim}{\theta})| < \varepsilon$, for any arbitrarily chosen $\varepsilon > 0$ if $|\hat{\underset{\sim}{\theta}}_k - \underset{\sim}{\theta}| < \gamma$ for $\underset{\sim}{\theta} \in I$. Hence,

$$\begin{aligned}\Pr\left(|p_n(\hat{\underset{\sim}{\theta}}_k) - p_n(\underset{\sim}{\theta})| < \varepsilon\right) &\geq \Pr\left(|\hat{\underset{\sim}{\theta}}_k - \underset{\sim}{\theta}| < \gamma\ ,\ \underset{\sim}{\theta} \in I\right) \\ &\geq \Pr\left(|\hat{\underset{\sim}{\theta}}_k - \underset{\sim}{\theta}| < \gamma\right) - \Pr\left(\underset{\sim}{\theta} \notin I\right) \\ &\geq 1 - \eta\end{aligned}$$

for $k > K$.

Theorem 2.2 – If the conditions of Theorem 2.1 are satisfied and since

$$\underset{n\to\infty}{\text{p lim}}\ p_n(\underset{\sim}{\theta}) = g(\underset{\sim}{\theta}) \tag{2.13}$$

then

$$\underset{\substack{n\to\infty \\ k\to\infty}}{\text{p lim}}\; p_n(\hat{\underset{\sim}{\theta}}_k) = g(\underset{\sim}{\theta}) \quad . \tag{2.14}$$

Proof — By Theorem 2.1, given any positive constants γ and ε , there exists an integer K such that for $k > K$

$$\Pr\left(|p_n(\hat{\underset{\sim}{\theta}}_k) - p_n(\underset{\sim}{\theta})| < \frac{\varepsilon}{2}\right) \geq (1 - \gamma)^{1/2} \quad ;$$

and since $p_n(\underset{\sim}{\theta})$ is consistent for $g(\underset{\sim}{\theta})$, there exists an integer N such that for $n > N$

$$\Pr\left(|p_n(\underset{\sim}{\theta}) - g(\underset{\sim}{\theta})| < \frac{\varepsilon}{2}\right) \geq (1 - \gamma)^{1/2} \quad .$$

Now

$$|p_n(\hat{\underset{\sim}{\theta}}_k) - g(\underset{\sim}{\theta})| \leq |p_n(\hat{\underset{\sim}{\theta}}_k) - p_n(\underset{\sim}{\theta})| + |p_n(\underset{\sim}{\theta}) - g(\underset{\sim}{\theta})| \quad .$$

Thus for $n > Z$ and $k > Z$ where $Z = \max(N,K)$,

$$\begin{aligned} &\Pr\left(|p_n(\hat{\underset{\sim}{\theta}}_k) - g(\underset{\sim}{\theta})| < \varepsilon\right) \\ &\quad \geq \Pr\left(\left(|p_n(\hat{\underset{\sim}{\theta}}_k) - p_n(\underset{\sim}{\theta})| + |p_n(\underset{\sim}{\theta}) - g(\underset{\sim}{\theta})|\right) < \varepsilon\right) \\ &\quad \geq \Pr\left(|p_n(\hat{\underset{\sim}{\theta}}_k) - p_n(\underset{\sim}{\theta})| < \frac{\varepsilon}{2} \;,\; |p_n(\underset{\sim}{\theta}) - g(\underset{\sim}{\theta})| < \frac{\varepsilon}{2}\right) \end{aligned}$$

The events

$$|p_n(\hat{\underset{\sim}{\theta}}_k) - p_n(\underset{\sim}{\theta})| < \frac{\varepsilon}{2} \tag{2.15}$$

and

$$|p_n(\underset{\sim}{\theta}) - g(\underset{\sim}{\theta})| < \frac{\varepsilon}{2} \tag{2.16}$$

are independent since the occurrence of (2.15) depends only on the kernel of integration $f(\hat{\underset{\sim}{\theta}}_k|\underset{\sim}{\theta})$, and the occurrence of (2.16) depends only on the prior density $g(\underset{\sim}{\theta})$. These densities are clearly independent; hence

$$\Pr\left(|p_n(\hat{\underset{\sim}{\theta}}_k) - g(\underset{\sim}{\theta})| < \varepsilon\right)$$

$$\geq \Pr\left(|p_n(\hat{\underset{\sim}{\theta}}_k) - p_n(\underset{\sim}{\theta})| < \frac{\varepsilon}{2}\right) \Pr\left(|p_n(\underset{\sim}{\theta}) - g(\underset{\sim}{\theta})| < \frac{\varepsilon}{2}\right)$$

$$\geq (1 - \gamma)^{1/2} (1 - \gamma)^{1/2} = 1 - \gamma \quad .$$

2.4 Continuously Smooth Estimators

The limiting result in (2.14) is somewhat artificial in practical applications since small values of k and n are usually encountered. Nevertheless, this property does suggest the replacement of the prior density in the Bayes estimator (2.1) by $p_n(\hat{\underset{\sim}{\theta}}_k)$ when considered as a function of $\underset{\sim}{\theta}$. The *continuously smooth empirical Bayes estimator* for θ_j , the j*th* component of $\underset{\sim}{\theta}_n$, becomes

$$\tilde{\Theta}_D = \frac{\int_{\underset{\sim}{\Theta}} \theta_j f(\hat{\underset{\sim}{\theta}}_{k,n} | \underset{\sim}{\theta}) p_n^*(\underset{\sim}{\theta})\, d\underset{\sim}{\theta}}{\int_{\underset{\sim}{\Theta}} f(\hat{\underset{\sim}{\theta}}_{k,n} | \underset{\sim}{\theta}) p_n^*(\underset{\sim}{\theta})\, d\underset{\sim}{\theta}} , \quad j = 1,2,\cdots,s \tag{2.17}$$

where

$$p_n^*(\underset{\sim}{\theta}) = \frac{1}{nh^s} \sum_{i=1}^{n} W\left(\frac{\underset{\sim}{\theta} - \hat{\underset{\sim}{\theta}}_{k,i}}{h} \right) . \tag{2.18}$$

For simplicity we have not indexed $\tilde{\Theta}_D$ with a subscript j . Denote the vector having components given by (2.17) as $\tilde{\underset{\sim}{\theta}}_D$. In particular for $s = 1$, (2.17) takes the form

$$\tilde{\Theta}_D = \frac{\sum_{i=1}^{n} \int_{\Theta} \theta f(\hat{\theta}_{k,n} | \theta) \left[\frac{\sin\left(\frac{\theta - \hat{\theta}_{k,i}}{2h} \right)}{\left(\frac{\theta - \hat{\theta}_{k,i}}{2h} \right)} \right]^2 d\theta}{\sum_{i=1}^{n} \int_{\Theta} f(\hat{\theta}_{k,n} | \theta) \left[\frac{\sin\left(\frac{\theta - \hat{\theta}_{k,i}}{2h} \right)}{\left(\frac{\theta - \hat{\theta}_{k,i}}{2h} \right)} \right]^2 d\theta} \tag{2.19}$$

where h is given by (2.8). The subscript D is a notational convenience and denotes the particular distribution of the estimator $\hat{\underset{\sim}{\theta}}_k$ when given $\underset{\sim}{\theta}$. For example, if the kernel of integration is normal then (2.17) would be represented as $\tilde{\theta}_N$.

In practice the actual range of $\underset{\sim}{\theta}$ will generally be unknown. This can be resolved satisfactorily by taking the region of integration in (2.17) to be the observed range of previous estimates. Thus it is necessary only to order successively the sequence of estimates given in (2.2), from which the range is easily calculated. The success of this approximation will be demonstrated in the remaining chapters.

Occasionally for certain families of distributions, no sufficient statistic exists for estimating $\underset{\sim}{\theta}$. In such cases some statistic $\hat{\underset{\sim}{\theta}}_k$ may exist which can be used in the formulation of $\tilde{\underset{\sim}{\theta}}_D$. This, of course, represents a further degree of approximation since $E(\underset{\sim}{\theta}|\underset{\sim}{x}) \neq E(\underset{\sim}{\theta}|\hat{\underset{\sim}{\theta}}_k)$. The estimator $\tilde{\underset{\sim}{\theta}}_D$, however, may continue to produce "good" results. This situation will occur in Chapter V and Chapter VI, and the continued success of $\tilde{\underset{\sim}{\theta}}_D$ will be witnessed.

The ability of the estimator $\tilde{\underset{\sim}{\theta}}_D$ to provide "good" results depends on the accuracy of several approximations, the most significant being the accuracy with which the density estimator $p_n^*(\underset{\sim}{\theta})$ represents the prior density $g(\underset{\sim}{\theta})$. If $\tilde{\underset{\sim}{\theta}}_D$ provides "better" estimates for the previous realizations of $\underset{\sim}{\theta}$ than the elements of sequence (2.2), then iteration of $\tilde{\underset{\sim}{\theta}}_D$ may achieve still further improvement. In each iteration the density estimator for the prior density function is based on the sequence of previous estimates $\tilde{\underset{\sim}{\theta}}_{D,i}$ $(i = 2,3,\cdots,n)$. When $i = 1$, $\tilde{\underset{\sim}{\theta}}_{D,1}$, is defined to be $\hat{\underset{\sim}{\theta}}_{k,1}$. Results from such iterations as well as further discussion are reported in the remaining chapters.

2.5 Marginal Variance Correction

In the construction of the continuously smooth empirical Bayes estimator $\tilde{\underset{\sim}{\theta}}_D$, it has been suggested that a consistent density estimator be used to represent the prior density. It would, therefore, seem desirable to base this density estimator on a sequence of values

$$\underset{\sim}{\theta}^*_{k,1},\ \underset{\sim}{\theta}^*_{k,2},\ \cdots,\ \underset{\sim}{\theta}^*_{k,n} \tag{2.20}$$

whose marginal distribution has mean and variance approximately equal to those of the prior density.

For finite samples of size k , the mean and variance of the marginal distribution of $\hat{\underset{\sim}{\theta}}_k$ are generally not equivalent to those of the prior distribution. A linear transformation, however, can be constructed on the elements of sequence (2.2) to provide a new sequence of values having a mean and variance which are approximately equal to the prior mean and variance. To illustrate this procedure when $s = 1$, consider the kernel of integration to be

$$\text{distr.}(\hat{\theta}_k \mid \theta) = N\left(\theta, \frac{\theta^2}{kc}\right) \tag{2.21}$$

where c is a known constant and k is the sample size. This procedure is quite general, and the normal distribution is only used for purposes of illustration. If the relations of conditional probability

$$E(\hat{\theta}_k) = E_\Theta\left[E(\hat{\theta}_k \mid \theta)\right] \tag{2.22}$$

and

$$\text{Var}(\hat{\theta}_k) = E_\Theta\left[\text{Var}(\hat{\theta}_k \mid \theta)\right] + \text{Var}_\Theta\left[E(\hat{\theta}_k \mid \theta)\right] \tag{2.23}$$

are applied to (2.21), then the mean and variance of the marginal distribution of $\hat{\theta}_k$ become

$$E(\hat{\theta}_k) = E(\theta)$$

and

$$\mathrm{Var}(\hat{\theta}_k) = E\left(\frac{\theta^2}{kc}\right) + \mathrm{Var}(\theta)$$

$$= \frac{1}{kc} E(\theta^2) + \mathrm{Var}(\theta)$$

$$= \frac{1}{kc}\left(\mathrm{Var}(\theta) + E^2(\theta)\right) + \mathrm{Var}(\theta)$$

$$= \frac{1+kc}{kc}\mathrm{Var}(\theta) + \frac{1}{kc}E^2(\theta) \quad . \qquad (2.24)$$

respectively. Thus for finite k, the mean of $\hat{\theta}_k$ unconditional on θ will be equal to the prior mean, but the variance of $\hat{\theta}_k$ overestimates the prior variance by the amount

$$\frac{1}{kc}\left[\mathrm{Var}(\theta) + E^2(\theta)\right] \quad .$$

In the light of this observation, consider the transformation of $\hat{\theta}_k$ given by

$$\theta_k^* = a_1\left(\hat{\theta}_k - E(\hat{\theta}_k)\right) + E(\theta) \qquad (2.25)$$

where a_1 is a constant to be determined. Now

$$E(\theta_k^*) = a_1 E(\hat{\theta}_k) - a_1 E(\hat{\theta}_k) + E(\theta) = E(\theta)$$

and

$$\mathrm{Var}(\theta_k^*) = \mathrm{Var}(a_1 \hat{\theta}_k)$$

$$= a_1^2 \, \mathrm{Var}(\hat{\theta}_k) \quad . \qquad (2.26)$$

Substituting (2.24) into (2.26) we obtain

$$\mathrm{Var}(\theta_k^*) = a_1^2 \left[\left(\frac{1+kc}{kc} \right) \mathrm{Var}(\theta) + \frac{1}{kc} E^2(\theta) \right]$$

$$= a_1^2 \left[\frac{(1+kc)\ \mathrm{Var}(\theta) + E^2(\theta)}{kc} \right] \quad . \qquad (2.27)$$

Hence by defining

$$a_1 = \left(\frac{kc\ \mathrm{Var}(\theta)}{(1+kc)\ \mathrm{Var}(\theta) + E^2(\theta)} \right)^{1/2} \qquad (2.28)$$

we obtain

$$\mathrm{Var}(\theta_k^*) = \mathrm{Var}(\theta)$$

the desired result.

If the mean and variance of the prior distribution are known, then a_1 can be exactly determined. In practice, however, these quantities will generally remain unknown; hence estimates for these quantities must be obtained. Since $E(\theta) = E(\hat{\theta}_k)$, the prior mean can be estimated by the sample mean

$$\bar{\theta}_n = \sum_{i=1}^{n} \frac{\hat{\theta}_{k,i}}{n} . \tag{2.29}$$

Estimating $\mathrm{Var}(\hat{\theta}_k)$ by the sample variance

$$s_n^2 = \frac{\sum_{i=1}^{n}\left(\hat{\theta}_{k,i} - \bar{\theta}_n\right)^2}{n-1} \tag{2.30}$$

and substituting (2.29) and (2.30) into (2.24), we have

$$s_n^2 = \frac{1+kc}{kc}\,\mathrm{Var}(\theta) + \frac{1}{kc}\,\bar{\theta}_n^2 . \tag{2.31}$$

Solving (2.31) for $\mathrm{Var}(\theta)$, we obtain

$$\widehat{\mathrm{Var}}(\theta) = \frac{kc}{1+kc}\,s_n^2 - \frac{1}{1+kc}\,\bar{\theta}_n^2 \tag{2.32}$$

as an estimate of the prior variance. Therefore when both the mean and variance of the prior distribution are unknown, $\bar{\theta}_n$ and $\widehat{\mathrm{Var}}(\theta)$ can be substituted for $E(\theta)$ and $\mathrm{Var}(\theta)$ in (2.28) giving

$$a_1 = \left[\frac{1}{1+kc} \left(kc - \frac{\overline{\theta}_n^2}{s_n^2} \right) \right]^{1/2} \tag{2.33}$$

The transformation Θ_k^* defined by (2.25) is sufficient for estimating Θ since it is obtained by a one-to-one transformation on the sufficient statistic $\hat{\Theta}_k$. To show that it is also consistent, consider

$$\underset{k\to\infty}{\text{p lim}}\ \Theta_k^* = \underset{k\to\infty}{\text{p lim}} \left[a_1 \hat{\Theta}_k - a_1 E(\hat{\theta}_k) + E(\theta) \right] \tag{2.34}$$

Now it is easily shown that both representations for a_1 , (2.28) and (2.33), have the property that

$$\lim_{k\to\infty} a_1 = 1 .$$

Thus (2.34) becomes

$$\underset{k\to\infty}{\text{p lim}}\ \Theta_k^* = \underset{k\to\infty}{\text{p lim}}\ \hat{\Theta}_k = \Theta \tag{2.35}$$

and Θ_k^* is therefore a consistent estimator for Θ .

Hence by Theorems 2.1 and 2.2, a density estimator based on sequence (2.20) has the property that

$$\underset{\substack{n\to\infty \\ k\to\infty}}{\mathrm{p\ lim}}\ p_n(\theta_k^*) = g(\theta) \quad . \tag{2.36}$$

CHAPTER III

ESTIMATION IN THE POISSON DISTRIBUTION

The Poisson distribution has played a significant role in the development of empirical Bayes techniques. Robbins [26] first introduced the empirical Bayes approach with the Poisson distribution. Recently, new empirical Bayes methods have been illustrated using this distribution. In keeping with the historical significance of the Poisson distribution, the usefulness of the continuously smooth empirical Bayes method will first be illustrated with this distribution.

Maximum-likelihood estimation is chosen as the classical method of estimation, and results from Monte Carlo simulations are reported, which show that the smooth estimators have smaller mean-squared errors than the maximum-likelihood estimators. These results are reported for small sample sizes and few past experiences.

3.1 Maximum-Likelihood Estimation

Assume that the conditional distribution of X, given any value $\theta > 0$, is Poisson with probability mass function

$$f(x|\theta) = \frac{e^{-\theta}\theta^{x}}{x!} \qquad (x = 0,1,\cdots;\ \theta > 0) \quad . \tag{3.1}$$

Consider a random sample of k observations from (3.1). The likelihood function of this sample is

$$L(\underset{\sim}{x}|\theta) = \frac{e^{-k\theta}\theta^{\sum_{i=1}^{k} x_i}}{\prod_{i=1}^{k} x_i!} . \tag{3.2}$$

Since the Poisson distribution satisfies the necessary regularity conditions given in section 1.7, the maximum-likelihood estimator for θ can be found by the solution of equation (1.22). Therefore taking the logarithm of (3.2) and differentiating with respect to θ, we have

$$\frac{d \log L(\underset{\sim}{x}|\theta)}{d\theta} = -k + \theta^{-1}\sum_{i=1}^{k} x_i . \tag{3.3}$$

Equating (3.3) to zero and solving for θ, we obtain the maximum-likelihood estimator

$$\hat{\theta}_k = \sum_{i=1}^{k} \frac{x_i}{k} . \tag{3.4}$$

Observation of equation (3.2) reveals that the likelihood function can be expressed as the product

$$L(\underset{\sim}{x}|\theta) = Q(t|\theta)\ C(\underset{\sim}{x})$$

where

$$C(\underset{\sim}{x}) = \frac{1}{\prod_{i=1}^{k} x_i!}$$

and

$$Q(t|\theta) = \theta^t e^{-k\theta} \quad .$$

Thus by the factorization criterion [23]

$$T = \sum_{i=1}^{k} x_i$$

is a sufficient statistic for Θ . The maximum-likelihood estimator $\hat{\Theta}_k$ can clearly be obtained by a one-to-one transformation on T and therefore represents a sufficient estimator for θ .

It is evident that given any $\theta > 0$, the only possible values for $\hat{\Theta}_k$ are $0, 1/k, 2/k, \cdots$. For a particular value t/k , the values $\underset{\sim}{x} = (x_1, x_2, \cdots, x_k)$ must be such that $\sum x_i = t$; hence the probability $f(t/k|\theta)$, that $\hat{\Theta}_k$ takes on the values t/k , is obtained by summing (3.2) over all sets $\underset{\sim}{x}$ so that $\sum x_i = t$.
That is,

$$f(\tfrac{t}{k}|\theta) = \sum_A \frac{e^{-k\theta}\theta^t}{\prod_{i=1}^{k} x_i!}$$

$$= e^{-k\theta}\theta^t \sum_A \frac{1}{\prod_{i=1}^{k} x_i!} \tag{3.5}$$

where A represents all possible sets $\underset{\sim}{x}$ such that $\sum x_i = t$. The multinomial theorem states that

$$(y_1 + y_2 + \cdots + y_k)^t = \sum_A t! \prod_{i=1}^{k} \frac{y_i^{x_i}}{x_i!} \quad ; \tag{3.6}$$

therefore setting each $y_i = 1$, we have

$$\frac{k^t}{t!} = \sum_A \frac{1}{\prod_{i=1}^{k} x_i!} \quad ,$$

and from (3.5)

$$f(\tfrac{t}{k}|\theta) = \frac{e^{-k\theta}(\theta k)^t}{t!} \quad ; \quad (\hat{\theta}_k = \tfrac{t}{k} = 0, \tfrac{1}{k}, \cdots) \quad . \tag{3.7}$$

Writing (3.7) explicitly as a function of the maximum-likelihood estimator $\hat{\theta}_k$, we obtain

$$f(\hat{\theta}_k|\theta) = \frac{e^{-k\theta}(k\theta)^{k\hat{\theta}_k}}{(k\hat{\theta}_k)!} \tag{3.8}$$

Since there is a one-to-one correspondence between $\hat{\theta}_k = T/k$ and $T = \sum x_i$, the probability mass function of T is

$$f(t|\theta) = \frac{e^{-k\theta}(k\theta)^t}{t!} , \quad (t = 0,1,2,\cdots) . \tag{3.9}$$

Hence

$$f(\hat{\theta}_k|\theta) = f(t|\theta) \tag{3.10}$$

which is a Poisson mass function with mean and variance given by $k\theta$.

3.2 Smooth Empirical Bayes Estimators for Θ

Assume that an unobservable random parameter θ occurs according to the unknown density function $g(\theta)$. When θ is realized, an observable random vector $\underset{\sim}{X}$ from (3.1) occurs, and a maximum-likelihood estimate of θ is formed using (3.4). Now if this situation occurs repeatedly, then the sequence

$$\hat{\theta}_{k,1}, \hat{\theta}_{k,2}, \cdots, \hat{\theta}_{k,n} \tag{3.11}$$

of maximum-likelihood estimates can be used to form a marginal density approximation $p_n(\hat{\theta}_k)$, given by (2.7), for $p(\hat{\theta}_k)$. Convergence in probability of $p_n(\hat{\theta}_k)$ to the unknown prior density function $g(\theta)$ is assured by Theorem 2.2. Therefore when (3.8) is used as the kernel of integration, $p_n(\hat{\theta}_k)$ can be treated as a function of Θ, and a continuously smooth empirical Bayes estimator for θ_n, the *nth* realization from $g(\theta)$, becomes

$$\tilde{\Theta}_p = \frac{\displaystyle\sum_{i=1}^{n} \int_{\hat{\theta}_{k,(1)}}^{\hat{\theta}_{k,(n)}} \theta e^{-k\theta}\theta^{k\hat{\theta}_{k,n}} \left[\frac{\sin\left(\frac{\theta-\hat{\theta}_{k,i}}{2h}\right)}{\left(\frac{\theta-\hat{\theta}_{k,i}}{2h}\right)} \right]^2 d\theta}{\displaystyle\sum_{i=1}^{n} \int_{\hat{\theta}_{k,(1)}}^{\hat{\theta}_{k,(n)}} e^{-k\theta}\theta^{k\hat{\theta}_{k,n}} \left[\frac{\sin\left(\frac{\theta-\hat{\theta}_{k,i}}{2h}\right)}{\left(\frac{\theta-\hat{\theta}_{k,i}}{2h}\right)} \right]^2 d\theta} . \tag{3.12}$$

The limits of integration in (3.12) $\hat{\theta}_{k,(1)}$ and $\hat{\theta}_{k,(n)}$ are the respective minimum and maximum values of sequence (3.11).

3.3 A Smooth Empirical Bayes Estimator for Θ Corrected for Variance

Although

$$\underset{\substack{n\to\infty \\ k\to\infty}}{\text{p lim}}\; p_n(\hat{\theta}_k) = g(\theta) ,$$

in practical application both n and k will remain finite. The marginal density estimator $p_n(\hat{\theta}_k)$ may therefore represent a poor approximation to the prior density. As demonstrated in section 2.5, when the mean and/or the variance of the marginal distribution of $\hat{\theta}_k$ are not equivalent to those of the prior distribution, the prior density approximation can often be improved.

Performing the transformation

$$\theta_k^* = a_1\left(\hat{\theta}_k - E(\hat{\theta}_k)\right) + E(\theta) \tag{3.13}$$

on each of the maximum-likelihood estimates of sequence (3.11), we obtain a sequence of values

$$\theta_{k,1}^*,\ \theta_{k,2}^*,\ \cdots,\ \theta_{k,n}^* \ . \tag{3.14}$$

The marginal distribution of this sequence has a mean and a variance equivalent to those of the prior distribution. Based on this sequence, an approximation $p_n(\theta_k^*)$ to the marginal density $p(\theta_k^*)$ can be formed. When considered as a function of θ , $p_n(\theta_k^*)$ can be used to represent the prior density in the Bayes estimator $E(\theta|\hat{\theta}_k)$. Thus an alternative smooth estimate of θ_n , the present realization from $g(\theta)$, can be given.

If the relations of conditional probability (2.22) and (2.23) are used, then the mean and the variance for the marginal distribution of the maximum-likelihood estimator become

$$E(\hat{\theta}_k) = E(k\theta) = kE(\theta) \tag{3.15}$$

and

$$\begin{aligned} \mathrm{Var}(\hat{\theta}_k) &= E(k\theta) + \mathrm{Var}(k\theta) \\ &= kE(\theta) + k^2\,\mathrm{Var}(\theta) \end{aligned} \tag{3.16}$$

respectively. Since the mean and variance of the marginal distribution of $\hat{\theta}_k$ overestimate the mean and variance of the prior distribution, the transformation given by (3.13) can be applied.

To determine the constant a_1 in the transformation (3.13) so that $\mathrm{Var}(\theta_k^*) = \mathrm{Var}(\theta)$, consider the variance of θ_k^* :

$$\mathrm{Var}(\theta_k^*) = \mathrm{Var}(a_1\,\hat{\theta}_k) = a_1^2\,\mathrm{Var}(\hat{\theta}_k) \ . \tag{3.17}$$

Substituting (3.16) for $\mathrm{Var}(\hat{\theta}_k)$ in (3.17), we have

$$\mathrm{Var}(\theta_k^*) = a_1^2\Big(kE(\theta) + k^2\ \mathrm{Var}(\theta)\Big) \quad . \qquad (3.18)$$

Thus choosing

$$a_1 = \left[\frac{\mathrm{Var}(\theta)}{kE(\theta) + k^2\ \mathrm{Var}(\theta)}\right]^{1/2} , \qquad (3.19)$$

we obtain the desired result.

When the mean and the variance of Θ are known, a_1 can be exactly determined by (3.19). In practice however, these values are generally unknown and require estimation. Since $E(\hat{\theta}_k) = kE(\theta)$, the prior mean can be estimated by $\overline{\theta}_n/k$ where $\overline{\theta}_n$ is the sample mean given by (2.29). The sample variance s_n^2 given by (2.30) can be used to approximate $\mathrm{Var}(\hat{\theta}_k)$. Proper substitution of these quantities into (3.16) gives

$$\widehat{\mathrm{Var}(\theta)} = \frac{s_n^2 - \overline{\theta}_n}{k^2} \qquad (3.20)$$

as an estimate of the prior variance. Thus when $E(\theta)$ and $\mathrm{Var}(\theta)$ are unknown the constant a_1 can be approximated by

$$a_1 = \left(\frac{s_n^2 - \overline{\theta}_n}{k^2 s_n^2}\right)^{1/2} \quad . \qquad (3.21)$$

With the transformation θ_k^* completely determined, the marginal density approximation $p_n(\theta_k^*)$ for $p(\theta_k^*)$ can be formed. Substituting $p_n(\theta_k^*)$, when considered as a function of θ, for $g(\theta)$ in the Bayes estimator $E(\theta|\hat{\theta}_k)$, we obtain as an alternative smooth estimator for θ_n

$$\tilde{\theta}_{P,V} = \frac{\displaystyle\sum_{i=1}^{n} \int_{\theta^*_{k,(1)}}^{\theta^*_{k,(n)}} \theta e^{-k\theta}\theta^{k\hat{\theta}_{k,n}} \left[\frac{\sin\left(\frac{\theta-\theta^*_{k,i}}{2h}\right)}{\left(\frac{\theta-\theta^*_{k,i}}{2h}\right)}\right]^2 d\theta}{\displaystyle\sum_{i=1}^{n} \int_{\theta^*_{k,(1)}}^{\theta^*_{k,(n)}} e^{-k\theta}\theta^{k\hat{\theta}_{k,n}} \left[\frac{\sin\left(\frac{\theta-\theta^*_{k,i}}{2h}\right)}{\left(\frac{\theta-\theta^*_{k,i}}{2h}\right)}\right]^2 d\theta} \tag{3.22}$$

where $\theta^*_{k,(1)}$ and $\theta^*_{k,(n)}$ are the respective minimum and maximum values of sequence (3.14), and h is given by (2.8).

3.4 Iteration of the Smooth Estimators

Consider a sequence of smooth estimates from (3.12),

$$\tilde{\theta}_{P,1}, \tilde{\theta}_{P,2}, \ldots, \tilde{\theta}_{P,n} \quad , \tag{3.23}$$

obtained in each of n previous experiences. Based on this sequence, a more "precise" estimate of the realization θ_n can often be determined. The estimate is obtained by replacing the prior density in the Bayes estimator by the continuous density approximation

$$p_n(\theta) = \frac{1}{2\pi nh} \sum_{i=1}^{n} \left[\frac{\sin\left(\frac{\theta-\tilde{\theta}_{P,i}}{2h}\right)}{\left(\frac{\theta-\tilde{\theta}_{P,i}}{2h}\right)} \right]^2 \tag{3.24}$$

constructed with sequence (3.23). This represents an iteration of the smooth estimator $\tilde{\theta}_P$ and is denoted by $\tilde{\theta}_P'$ since the kernel of integration remains unchanged. A similar iteration of the estimator $\tilde{\theta}_{P,V}$ can be obtained and will be denoted by $\tilde{\theta}_{P,V}'$. Integration in each estimator is performed over the range of values obtained from the first iteration. For example, the region of integration for $\tilde{\theta}_P'$ is from $\min(\tilde{\theta}_{P,i})$ to $\max(\tilde{\theta}_{P,j})$ where $i,j = 1,2,\cdots,n$ and $i \neq j$. The smooth estimator $\tilde{\theta}_P$ is defined for $n > 1$; therefore for $i = 1$, we define $\tilde{\theta}_{P,1} = \hat{\theta}_{k,1}$.

3.5 Monte Carlo Simulation

To ascertain the usefulness of the continuously smooth empirical Bayes estimator $\tilde{\theta}_D$ as opposed to the maximum-likelihood estimator, Monte Carlo simulation was employed by means of a UNIVAC 1108 computer. The criterion for comparison was mean-squared error, and therefore the ratio

$$R = \frac{\text{empirical Bayes mean-squared error}}{\text{maximum-likelihood mean-squared error}} \tag{3.25}$$

was of interest. Here the symbol $\tilde{\theta}_D$ is used as a general notational device to represent any smooth estimator under consideration.

A value of θ was generated from a chosen prior distribution; then a random sample $x_1, x_2, \cdots, x_k$ of size k, corresponding to the realization of θ, was obtained from (3.1). The maximum-likelihood estimate $\hat{\theta}_k$ was found and its squared deviation $(\theta - \hat{\theta}_k)^2$ from the corresponding parameter θ was calculated. For the second experiment, a new value for θ was generated and the process repeated, obtaining $\hat{\theta}_k$ and its squared deviation. For this experiment, $\tilde{\theta}_D$ and its squared deviation $(\theta - \tilde{\theta}_D)^2$ from the corresponding realization of θ were also calculated. This was repeated twenty times, and each time, $\tilde{\theta}_D$ was calculated using the present $\hat{\theta}_k$ as well as all previous maximum-likelihood estimates. Five hundred repetitions of this run of twenty experiments were then made, and the averages of the squared deviations of $\hat{\theta}_k$ and $\tilde{\theta}_D$ were formed as estimates of $E(\theta - \hat{\theta}_k)^2$ and $E(\theta - \tilde{\theta}_D)^2$. Then the ratio R was calculated utilizing these estimated mean-squared errors. All numerical integrations were performed by means of the eleven-point Gauss quadrature formula. For details see Appendix A.

This procedure was repeated for all types of Pearson prior distributions with varying coefficients of skewness and kurtosis. As with Clemmer and Krutchkoff [3], Lemon and Krutchkoff [17], and Martz and Krutchkoff [20], the ratio R was observed to be significantly influenced by the prior distribution only through the value

$$Z = \frac{\mathrm{Var}^*(\hat{\theta}_k | \theta)}{\mathrm{Var}(\theta)} \tag{3.26}$$

where * indicates that E(θ), the prior mean of Θ, has been substituted for θ. In particular, for a random sample of k observations from (3.1), the maximum-likelihood estimator is distributed according to (3.8), and the value Z becomes

$$Z = \frac{kE(\theta)}{\mathrm{Var}(\theta)} \tag{3.27}$$

Since it was found that the only factors affecting the ratio R, apart from the number of experiences, are contained in (3.27), this quantity can be conveniently used to summarize and index a given situation.

It has been repeatedly observed in the Monte Carlo study that the sample size k has no effect on the ratio R when formed with $\tilde{\theta}_P$ or $\tilde{\theta}_{P,V}$. Therefore without loss of generality, k will be taken to be one. We will denote the ratio R when calculated with $\tilde{\theta}_P$ by R_P and when calculated with $\tilde{\theta}_{P,V}$ by $R_{P,V}$.

To support the claim that the smooth empirical Bayes estimators are indeed robust to the form of the prior distribution, the values of θ_i $(i = 1,2,\cdots,20)$ were generated from various distributions. For all types of Pearson prior distributions with varying coefficients of skewness and kurtosis, the ratio R has been observed to vary only slightly for a given value of n , providing the value of Z remains unchanged. This is illustrated in Figures 3-6 for a given value of $Z = 2.0$. The solid line in each figure represents the ratio $R_{P,V}$ calculated with the smooth estimator $\tilde{\theta}_{P,V}$; and the broken line represents the ratio R_P calculated with $\tilde{\theta}_P$. This representation will be used in the remaining figures of this chapter. The parameters of the prior distributions are designated as follows:

$E(\theta)$ = prior mean of Θ

$V(\theta)$ = prior variance of Θ

S = skewness

K = kurtosis

In Figure 3 the prior distribution is bell-shaped (skewed); in Figure 4, L-shaped; in Figure 5, J-shaped; and in Figure 6, U-shaped. These estimators are evidently rather insensitive to the form of the prior distribution as summarized by S and K; therefore, values of S and K will not be given for the remaining figures in this chapter.

Values of the ratios R_P and $R_{P,V}$ are plotted in Figures 7 and 8 respectively. These values are plotted for different values of $E(\theta)$ and $V(\theta)$, summarized by Z. The values of Z range from 0.5 to 5.0. We note that as Z increases, the values of the ratios $R_{P,V}$ and R_P tend to decrease. This phenomenon is best understood by considering the summary quantity Z, defined by (3.26). If $\mathrm{Var}(\hat{\theta}_k|\theta)$ is large as compared to $\mathrm{Var}(\theta)$, then the classical estimates of θ will vary widely. The smooth estimators, however, are capable of "detecting" this variation and use this information to obtain "better" estimates of θ. Conversely if $\mathrm{Var}(\hat{\theta}_k|\theta)$ is small as compared to $\mathrm{Var}(\theta)$, then the classical method would be expected to do quite well.

In this case there is a great deal of information within an experiment, and previous experiments contribute very little information about the parameter. We also notice that for a given number of experiences, the decrease of $R_{P,V}$ is more significant than that of R_P demonstrating the superiority of the smooth estimator $\tilde{\Theta}_{P,V}$ over the smooth estimator $\tilde{\Theta}_P$. This increase in mean-squared precision is easily explained. The prior density approximation used in $\tilde{\Theta}_{P,V}$ is based on a sequence of values having a marginal distribution whose mean and variance are approximately equivalent to those of the prior distribution. The prior density approximation used in $\tilde{\Theta}_P$, however, is based on a sequence of maximum-likelihood estimates having a marginal distribution whose mean and variance are known to overestimate those of the prior distribution. Hence we expect $\tilde{\Theta}_{P,V}$ to provide a significant increase in squared-error precision over $\tilde{\Theta}_P$ as seen by comparing the corresponding values of the ratios $R_{P,V}$ and R_P of Figures 7 and 8.

Figure 9 shows a typical result obtained when $\tilde{\Theta}_{P,V}$ is iterated as described in section 3.4. The ratio $R'_{P,V}$ formed with the iterated smooth estimator $\tilde{\Theta}'_{P,V}$ is denoted by the dotted line. We note that a second iteration of $\tilde{\Theta}_{P,V}$ slightly decreased any improvement obtained by $\tilde{\Theta}_{P,V}$. This decrease in precision is expected. The marginal density approximation used to represent the prior density in $\tilde{\Theta}_{P,V}$ is based on a sequence of estimates whose distribution has its mean and variance approximately equivalent to those of the prior distribution. The prior density approximation in $\tilde{\Theta}'_{P,V}$, however, is not known to have this property.

In Figure 10 results from a second iteration of $\tilde{\Theta}_P$ are given when the parameters of the prior distribution are identical to those used in Figure 9. The dotted line is used to represent the ratio R'_P based on the smooth estimator $\tilde{\Theta}'_P$. We note that in this case, iteration does improve the ratio R ; however, this improvement is not as significant as that obtained using $\tilde{\Theta}_{P,V}$ given in Figure 9. It is conjectured that the overestimation of the prior variance by the marginal distribution of $\hat{\Theta}_k$ has a far more significant effect on the prior density approximation than does the marginal variance of $\tilde{\Theta}_P$. In general, iteration of the smooth estimators is discouraged. While a second iteration of $\tilde{\Theta}_P$ does decrease the squared error, the decrease is not as significant as that obtained using $\tilde{\Theta}_{P,V}$. A second iteration of $\tilde{\Theta}_{P,V}$ usually increases the squared error.

In all cases considered in the Monte Carlo study, the estimator $\tilde{\Theta}_{P,V}$ provided consistent improvement over the maximum-likelihood estimator for two or more past experiences. It was also observed to be the most efficient of the smooth estimators. Since this improvement was uniform over a wide variety of Z values, we are confident that $\tilde{\Theta}_{P,V}$ can be used in any situation. However, when some idea of the prior mean and variance is known, Figure 11 can be used to obtain an indication of the amount of improvement $\tilde{\Theta}_{P,V}$ will provide over the maximum-likelihood estimator. In this figure, the ratio $R_{P,V}$ is plotted as a function of Z for a given number of experiences. Thus one can obtain some a priori idea of the improvement over maximum-likelihood that is likely to be obtained.

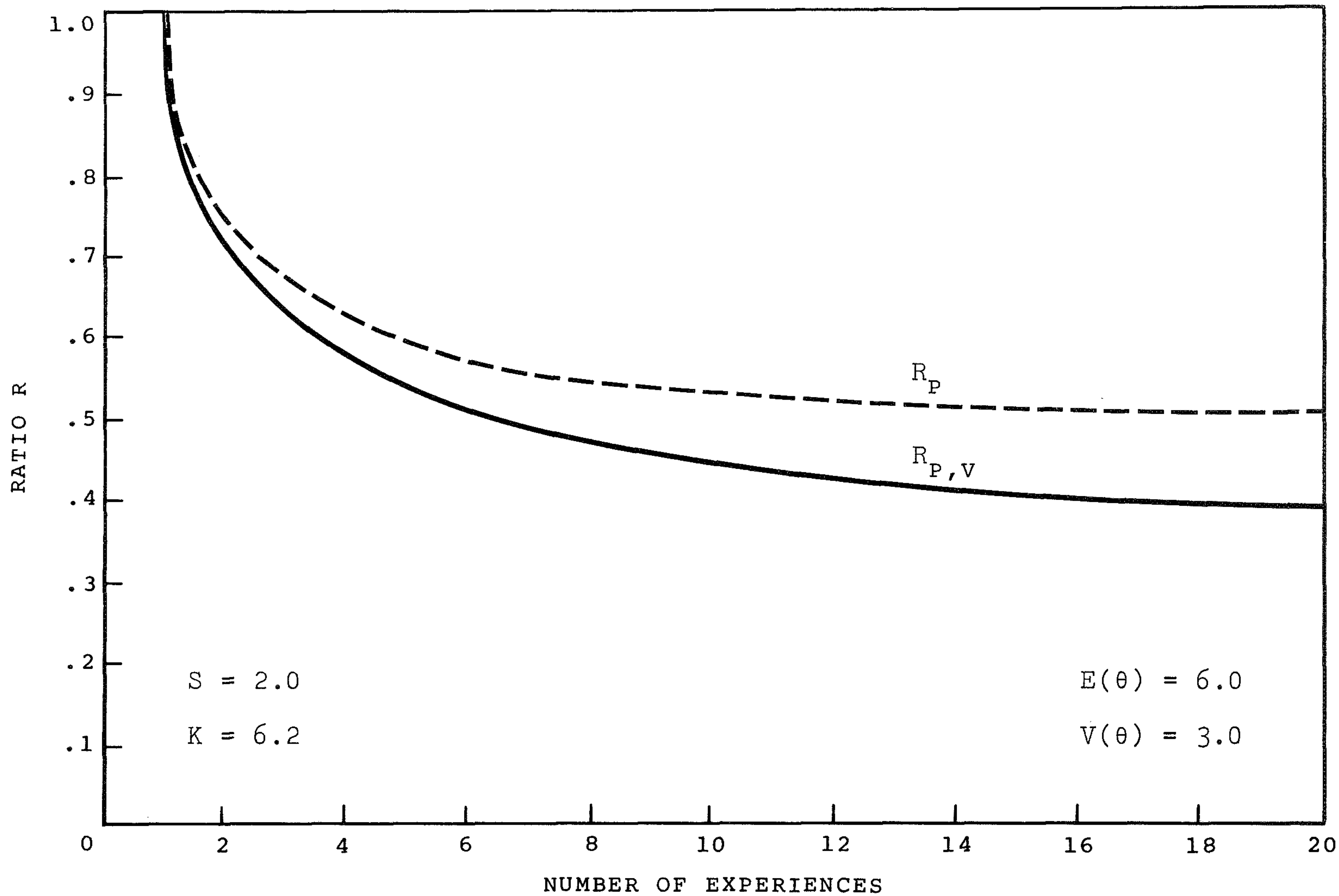

Figure 3. — R_P and $R_{P,V}$ for $Z = 2.0$; bell-shaped (skewed) prior.

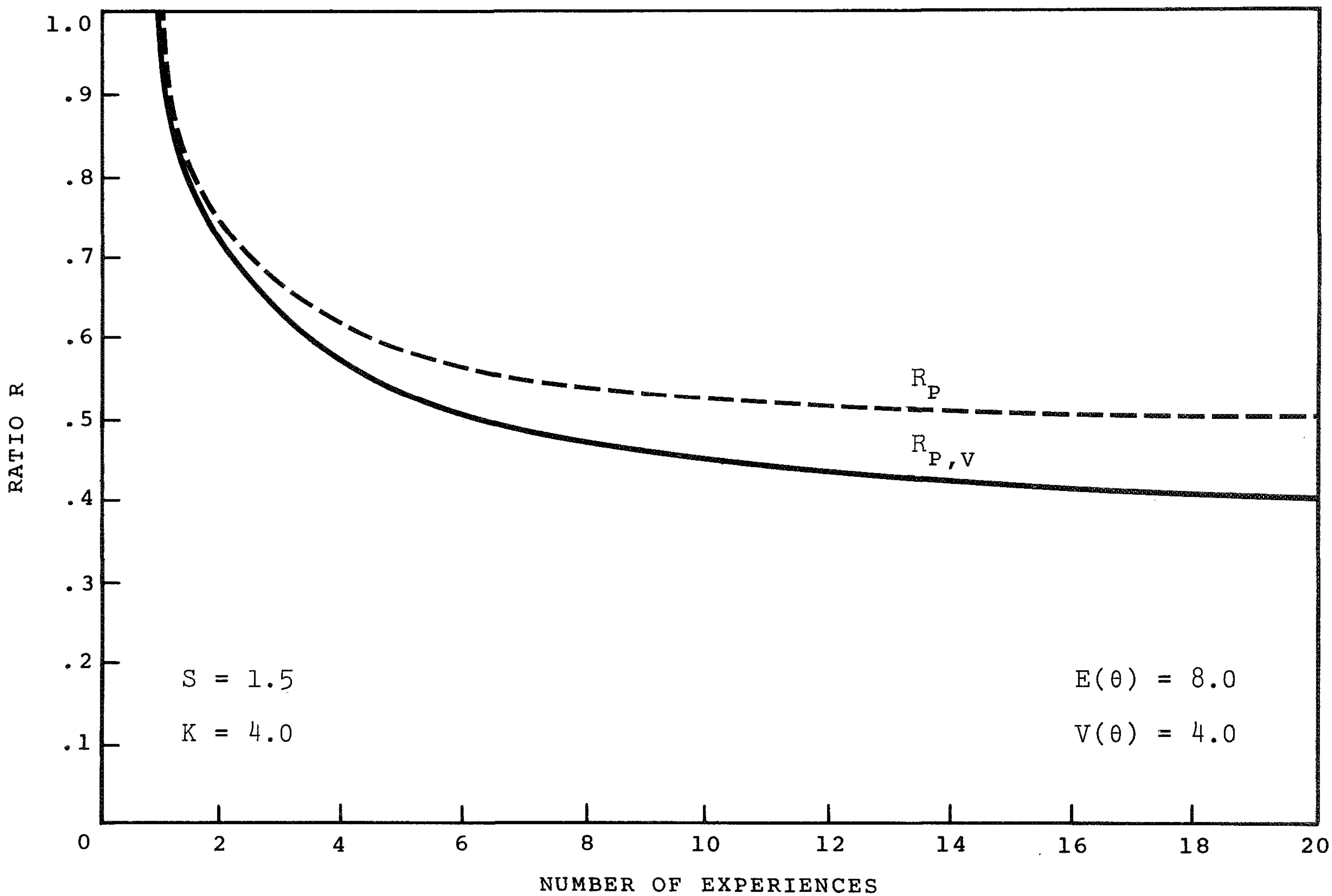

Figure 4. – R_P and $R_{P,V}$ for $Z = 2.0$; L-shaped prior.

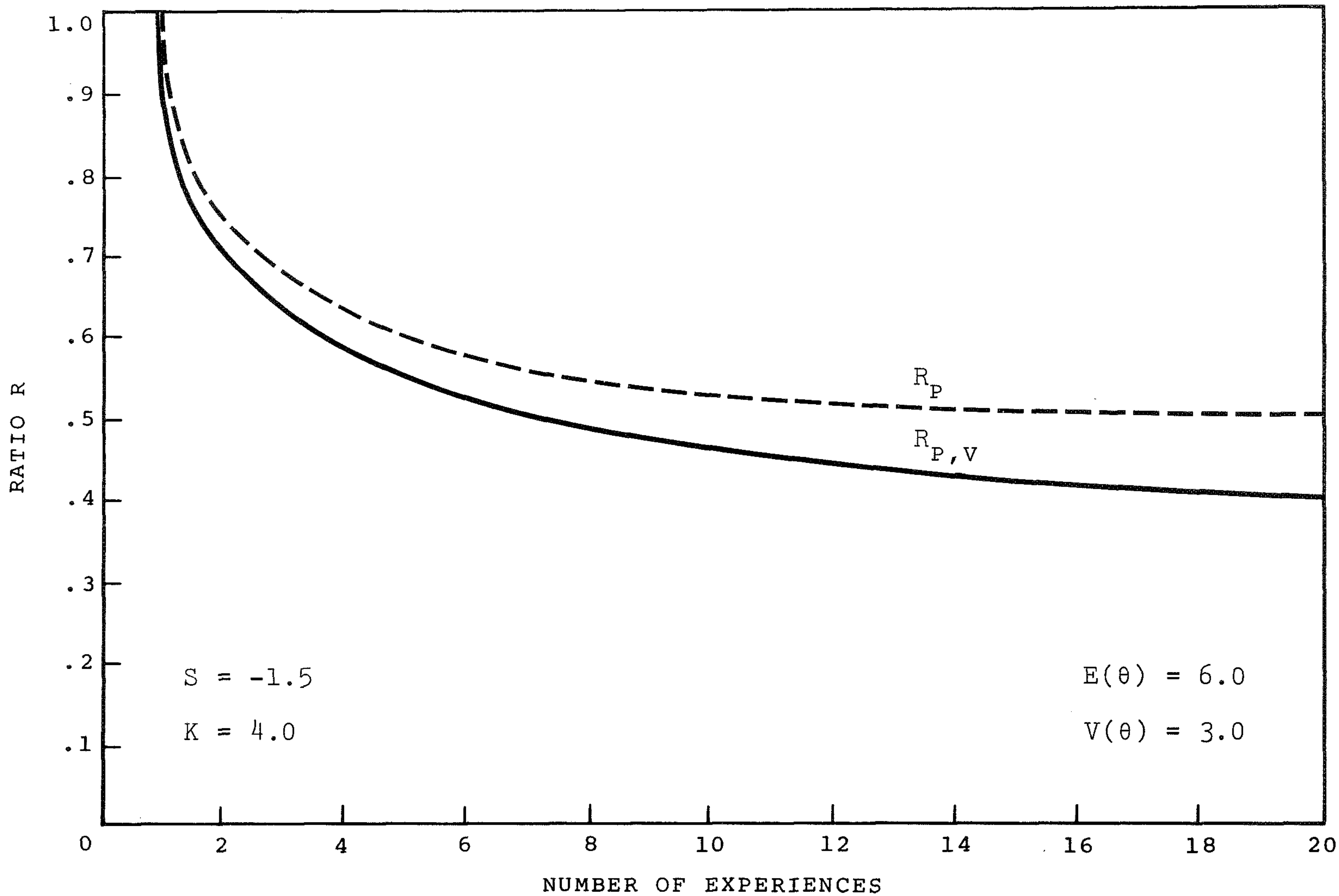

Figure 5. — R_P and $R_{P,V}$ for $Z = 2.0$; J-shaped prior.

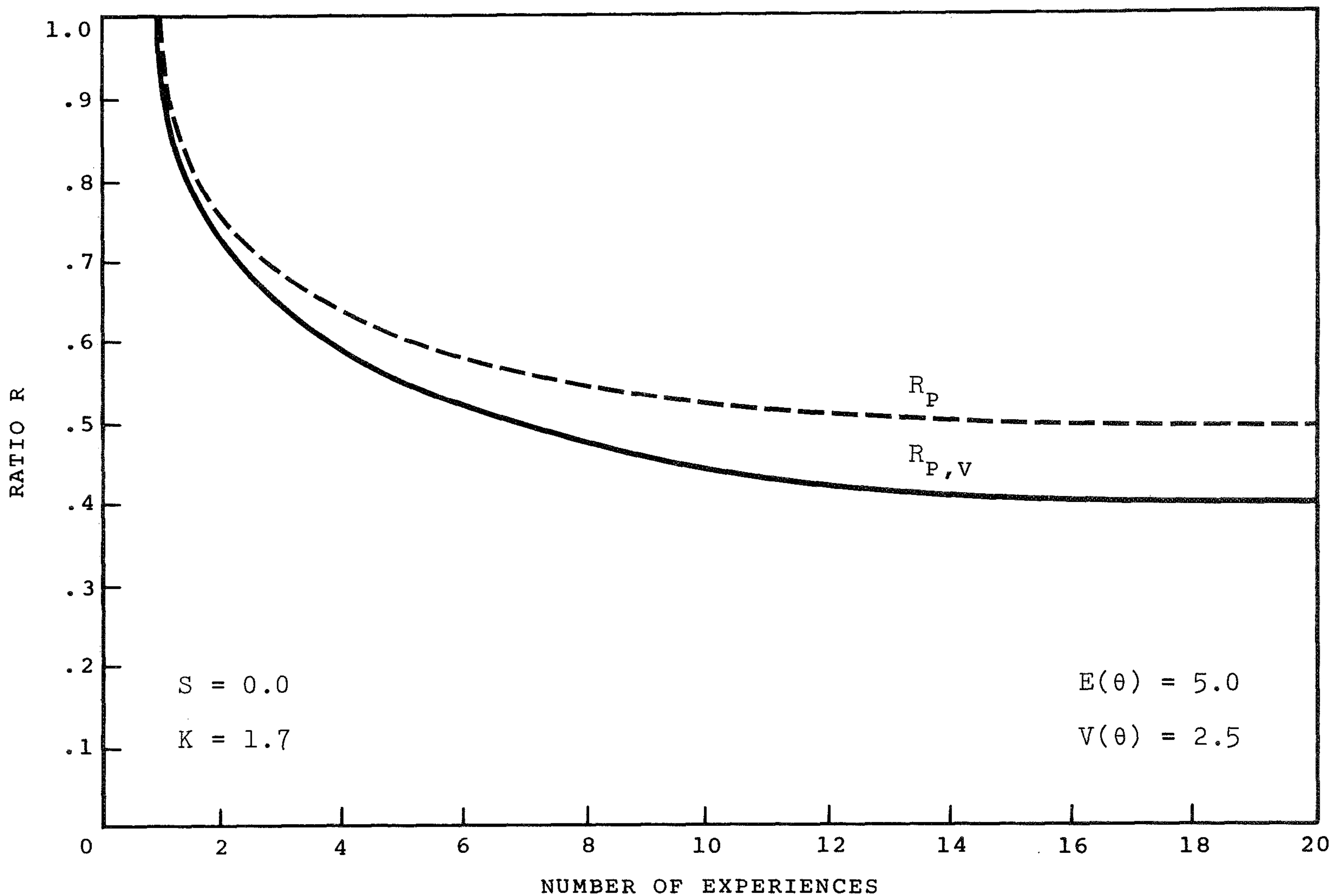

Figure 6. — R_P and $R_{P,V}$ for $Z = 2.0$; U-shaped prior.

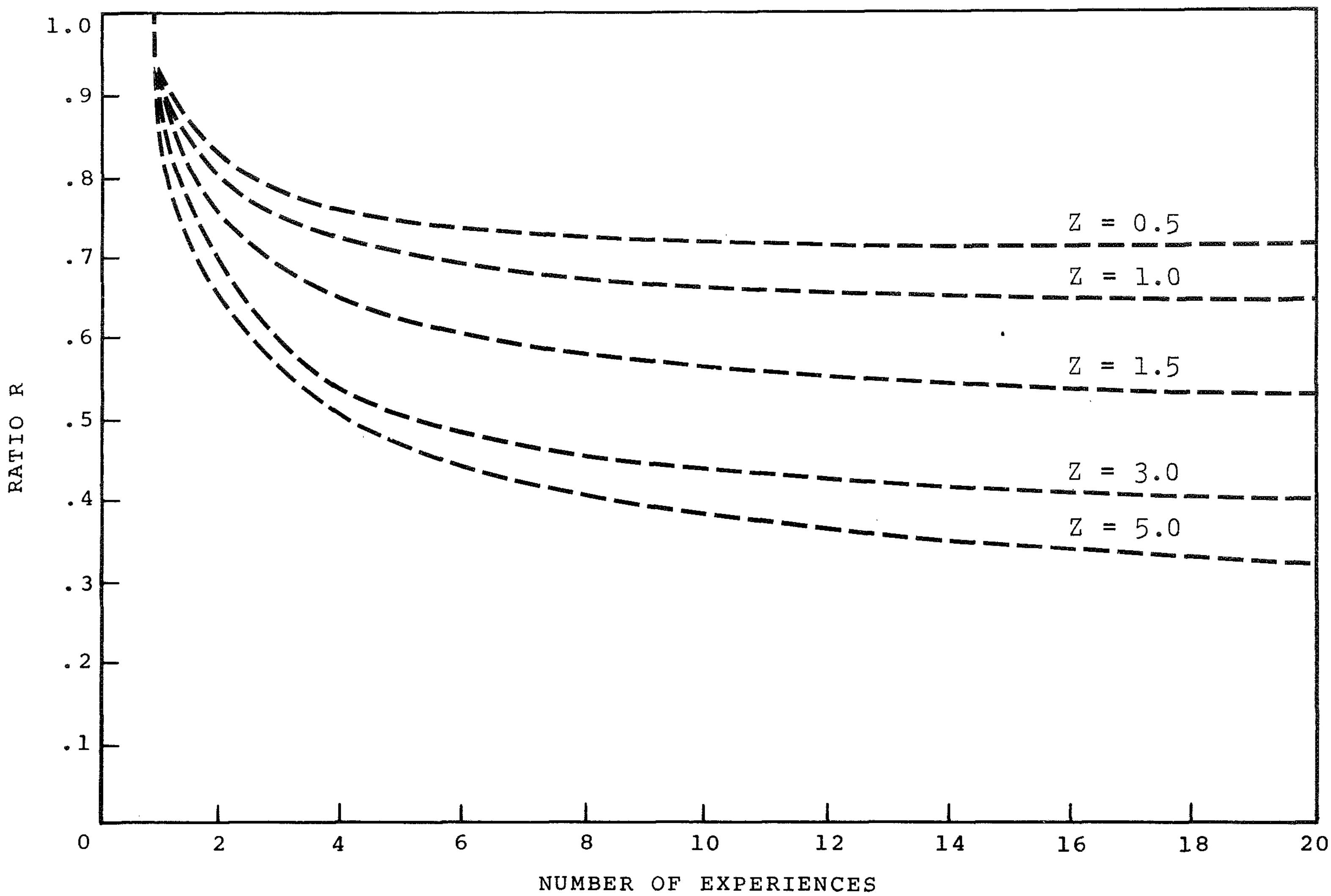

Figure 7. — R_p for several values of Z .

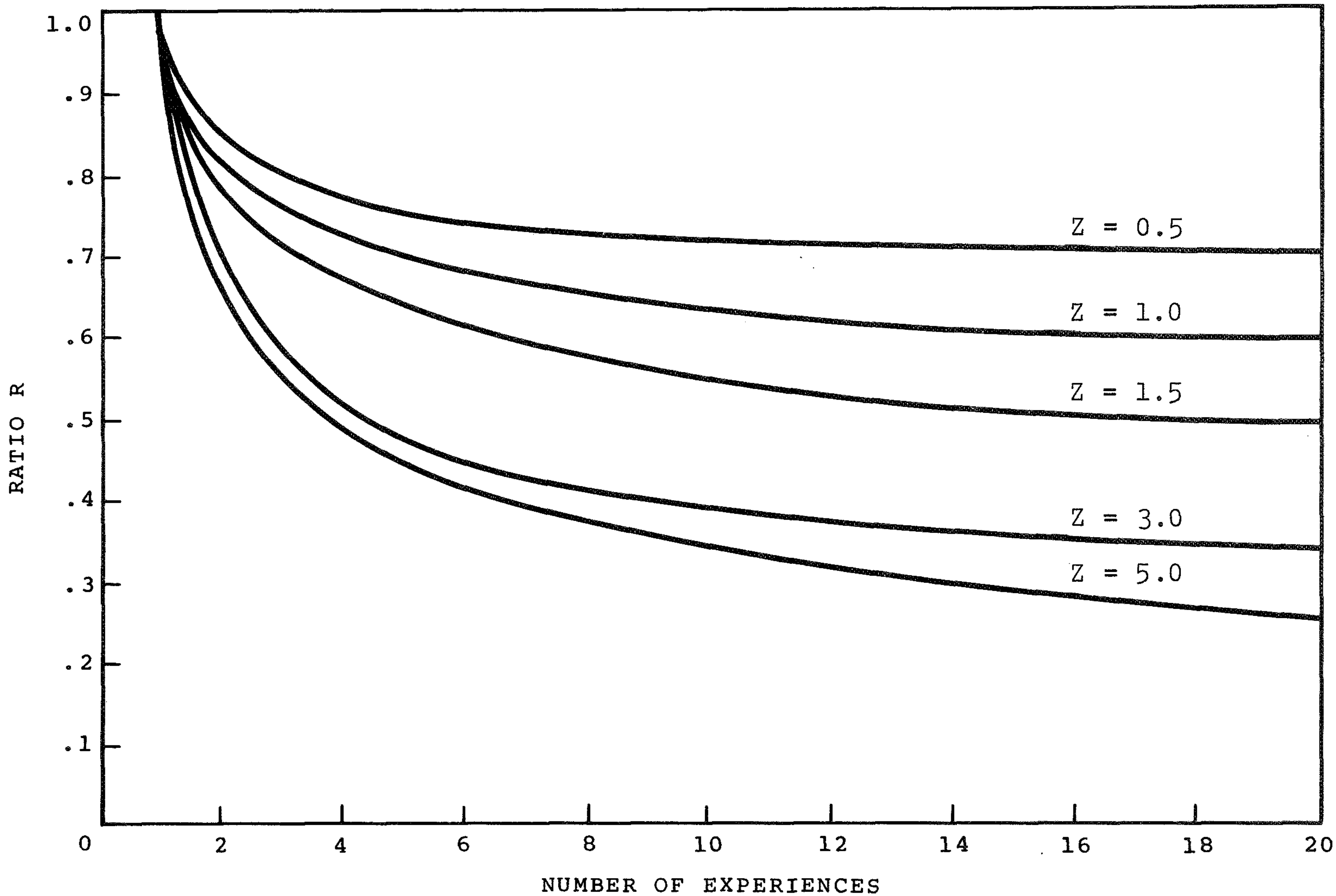

Figure 8. — $R_{P,V}$ for several values of Z .

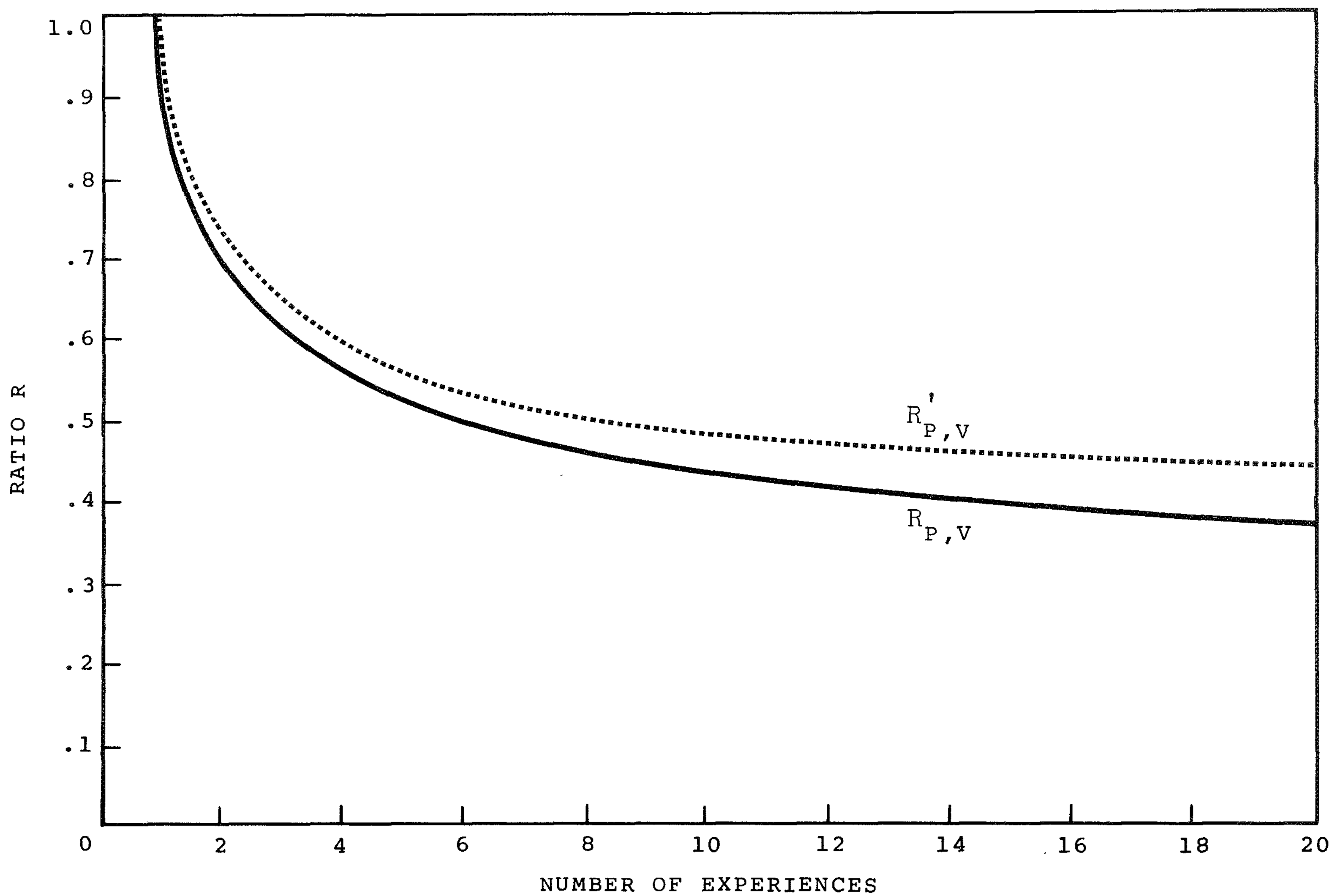

Figure 9. — $R_{P,V}$ vs $R'_{P,V}$ for $Z = 2.5$.

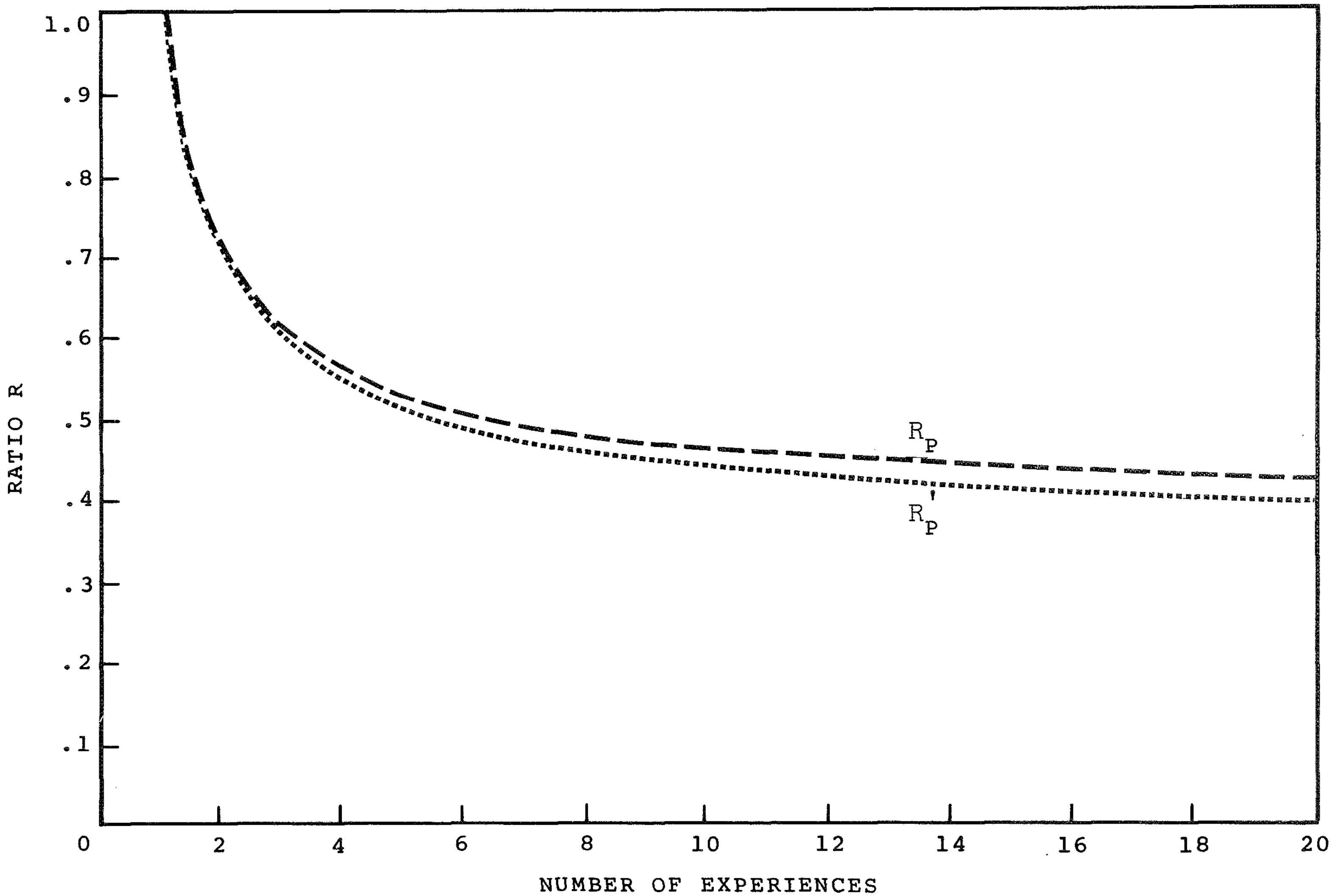

Figure 10. – R_p vs R_p' for $Z = 2.5$.

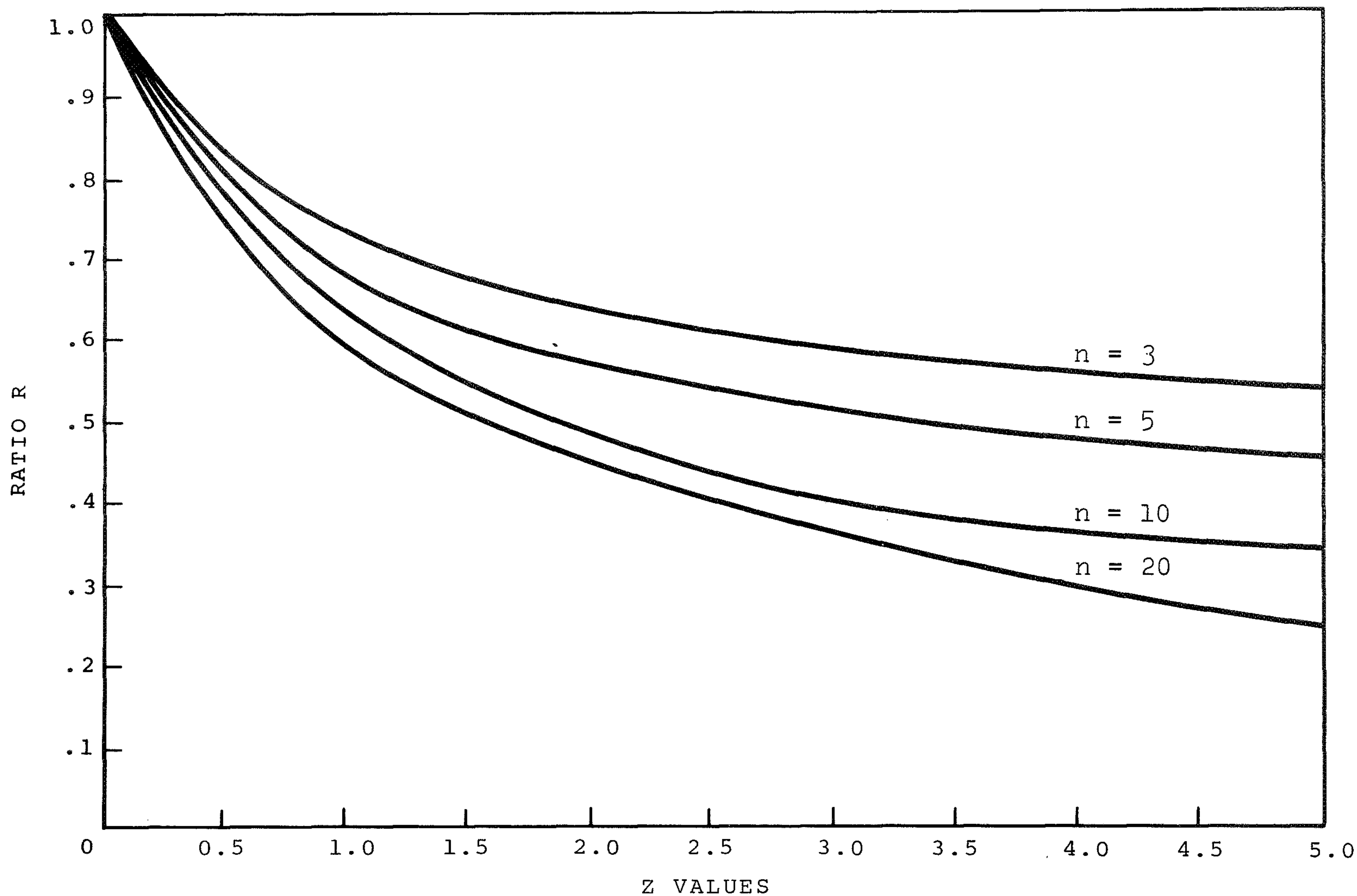

Figure 11. — $R_{P,V}$ as a function of Z for fixed number of experiences.

CHAPTER IV

ESTIMATION OF THE WEIBULL SCALE PARAMETER WITH KNOWN SHAPE PARAMETER

In this chapter continuously smooth empirical Bayes estimators are given for the scale parameter α in the two-parameter Weibull distribution. The scale parameter is assumed to vary randomly throughout a sequence of experiments and the shape parameter β is assumed to be known. It may, however, be different in each experiment. Results from Monte Carlo simulations are reported which show that even for small sample sizes and few experiments the smooth estimators have smaller mean-squared errors than the maximum-likelihood estimators.

4.1 Maximum-Likelihood Estimator for α

Let X be a random variable having a Weibull distribution with known shape parameter β. If the scale parameter α is a random variable, then the conditional density function of X is given by

$$f(x|\alpha) = \alpha\beta x^{\beta-1}e^{-\alpha x^{\beta}} , \qquad (x \geq 0;\ \alpha,\ \beta > 0) \quad . \ (4.1)$$

Consider a random sample of k observations from (4.1). The likelihood function of this sample is

$$L(\underset{\sim}{x}|\alpha) = (\alpha\beta)^k e^{-\alpha\left(\sum_{i=1}^{k} x_i^\beta\right)} \prod_{i=1}^{k} x_i^{\beta-1} , \qquad (4.2)$$

and by defining

$$T = \sum_{i=1}^{k} x_i^\beta \qquad (4.3)$$

can be expressed as the product

$$L(\underset{\sim}{x}|\alpha) = Q(t|\alpha)c(\underset{\sim}{x})$$

where

$$c(\underset{\sim}{x}) = \beta^k \prod_{i=1}^{k} x_i^{\beta-1}$$

and

$$Q(t|\alpha) = \alpha^k e^{-\alpha t} .$$

Thus, the factorization criterion [23] assures that T is a sufficient statistic for α .

The Weibull distribution satisfies the necessary regularity conditions given in section 1.7 to allow the maximum-likelihood estimator for α to be found by the solution of equation (1.22). Therefore, taking the logarithm of (4.2) and differentiating with respect to α, we have

$$\frac{d \log L(\underset{\sim}{x}|\alpha)}{d\alpha} = \frac{k}{\alpha} - \sum_{i=1}^{k} x_i^{\beta} \quad . \tag{4.4}$$

Equating (4.4) to zero and solving for α, we obtain the maximum-likelihood estimator

$$\hat{\alpha}_k = \frac{k}{\sum_{i=1}^{k} x_i^{\beta}} \quad . \tag{4.5}$$

This estimator can clearly be obtained by a one-to-one transformation on the sufficient statistic T; therefore the maximum-likelihood estimator $\hat{\alpha}_k$ provides a sufficient statistic for each realization of α. For ease of notation, the subscript k representing the sample size will be suppressed.

4.2 Distribution of the Maximum-Likelihood Estimator for α

According to (1.25), the asymptotic distribution of the maximum-likelihood estimator $\hat{\alpha}$ given the parameter α is normal with mean

$$E(\hat{\alpha}|\alpha) = \alpha \tag{4.6}$$

and variance given by

$$\mathrm{Var}(\hat{\alpha}|\alpha) = -\left[E_x\left(\frac{d^2 \log L(\underset{\sim}{x}|\alpha)}{d\alpha^2}\right)\right]^{-1} \tag{4.7}$$

If we differentiate (4.4) with respect to α and form the expectation with respect to X, then

$$E\left(\frac{d^2 \log L(\underset{\sim}{x}|\alpha)}{d\alpha^2}\right) = -\frac{k}{\alpha^2} \; . \tag{4.8}$$

Hence the variance becomes

$$\mathrm{Var}(\hat{\alpha}|\alpha) = \frac{\alpha^2}{k} \tag{4.9}$$

and

$$\text{distr.}(\hat{\alpha}|\alpha) = N\left(\alpha, \frac{\alpha^2}{k}\right) . \tag{4.10}$$

Since the maximum-likelihood estimator is representable in closed form, its conditional distribution can be found for finite values of k. This distribution will be obtained through a series of variable changes and will require the use of moment-generating functions.

Consider the distribution of the variable

$$Y = X^{\beta} \tag{4.11}$$

where the conditional density of X is given by (4.1). Since Y is either an increasing or decreasing function of X, depending on the fixed value β, the conditional density function of Y, say $h(y|\alpha)$, is given by the formula

$$h(y|\alpha) = f(x|\alpha)\left|\frac{dx}{dy}\right| \tag{4.12}$$

in which X is to be replaced by its value in terms of Y given in equation (4.1). This formula represents a simple change of variable technique which can be found in

Hoel [14]. Based on formula (4.12), the conditional density function of Y given α becomes the exponential density function

$$h(y|\alpha) = \alpha e^{-\alpha y} , \qquad (y \geq 0;\ \alpha > 0) . \qquad (4.13)$$

Corresponding to this density function, the moment-generating function is given by

$$M_Y(\theta) = \int_0^\infty \alpha e^{-(\alpha-\theta)y}\, dy$$

$$= \frac{1}{1 - \frac{\theta}{\alpha}} . \qquad (4.14)$$

Since each $x_i (i = 1,2,\cdots,k)$ is independently and identically distributed, the moment-generating function of T given α , $M_T(\theta)$, can be written as

$$M_T(\theta) = M_{x_1^\beta + x_2^\beta + \cdots + x_k^\beta}(\theta) = M_{x_1^\beta}(\theta)\, M_{x_2^\beta}(\theta) \cdots M_{x_k^\beta}(\theta) ,$$

and by (4.14) as

$$M_T(\theta) = \frac{1}{\left(1 - \frac{\theta}{\alpha}\right)^k} .$$

Now this function is precisely the moment-generating function for a gamma distribution with density function

$$f(t|\alpha) = \frac{e^{-\alpha t}(\alpha t)^{k-1}}{\Gamma(k)} \alpha \, , \qquad (t \geq 0;\ k,\ \alpha > 0) \quad ; \tag{4.15}$$

therefore (4.15) represents the conditional density function of T given α . Based on this result, the conditional distribution of the maximum-likelihood estimator is easily obtained. If the previously described change of variable procedure is followed, the maximum-likelihood estimator $\hat{\alpha}$ can be shown to have the conditional density function

$$f(\hat{\alpha}|\alpha) = \frac{\left(\frac{\alpha k}{\hat{\alpha}}\right)^{k+1} e^{-\frac{\alpha k}{\hat{\alpha}}}}{\Gamma(k)\alpha k} \quad . \tag{4.16}$$

This density function is recognizable as an inverted gamma density function with mean

$$E(\hat{\alpha}|\alpha) = \frac{\alpha\, k}{k-1} \tag{4.17}$$

and variance

$$\mathrm{Var}(\hat{\alpha}|\alpha) = \frac{(\alpha k)^2}{(k-1)^2(k-2)} \quad . \tag{4.18}$$

Thus for finite sample sizes, the maximum-likelihood estimator $\hat{\alpha}$ conditional on any α has the inverted gamma distribution given by (4.16).

4.3 Smooth Empirical Bayes Estimators for α

Assume that the data obtained in each of n previous experiments can be adequately described by a Weibull distribution with known shape parameter β_i. Further assume that between successive experiments, the scale parameter α varies randomly with the same but unknown prior density function $g(\alpha)$. If in each experiment a maximum-likelihood estimate $\hat{\alpha}_i (i = 1,2,\cdots,n)$ was obtained for each realization from $g(\alpha)$, then the sequence of sufficient and consistent estimates

$$\hat{\alpha}_1, \hat{\alpha}_2, \cdots, \hat{\alpha}_n \tag{4.19}$$

can be used to form the marginal density approximation

$$p_n(\hat{\alpha}) = \frac{1}{2\pi nh} \sum_{i=1}^{n} \left[\frac{\sin\left(\frac{\hat{\alpha}-\hat{\alpha}_i}{2h}\right)}{\left(\frac{\hat{\alpha}-\hat{\alpha}_i}{2h}\right)} \right]^2 \tag{4.20}$$

where h is given by (2.8).

Convergence in probability of (4.20) to the prior density $g(\alpha)$ as both the sample size k and the number of experiments n tend to infinity is assured by Theorem 2.2. Therefore when the kernel of integration is the inverted gamma density given by (4.16), $p_n(\hat{\alpha})$ can be considered as a function of α and a continuously smooth empirical Bayes estimator for α_n becomes

$$\tilde{\alpha}_G = \frac{\displaystyle\sum_{i=1}^{n} \int_{\hat{\alpha}_{(1)}}^{\hat{\alpha}_{(n)}} \left(\frac{\alpha k}{\hat{\alpha}_n}\right)^{k+1} e^{-\left(\frac{\alpha k}{\hat{\alpha}_n}\right)} \left[\frac{\sin\left(\frac{\alpha-\hat{\alpha}_i}{2h}\right)}{\left(\frac{\alpha-\hat{\alpha}_i}{2h}\right)}\right]^2 d\alpha}{\displaystyle\sum_{i=1}^{n} \int_{\hat{\alpha}_{(1)}}^{\hat{\alpha}_{(n)}} \frac{1}{\alpha}\left(\frac{\alpha k}{\hat{\alpha}_n}\right)^{k+1} e^{-\left(\frac{\alpha k}{\hat{\alpha}_n}\right)} \left[\frac{\sin\left(\frac{\alpha-\hat{\alpha}_i}{2h}\right)}{\left(\frac{\alpha-\hat{\alpha}_i}{2h}\right)}\right]^2 d\alpha} \tag{4.21}$$

where $\hat{\alpha}_{(1)}$ and $\hat{\alpha}_{(n)}$ are the respective minimum and maximum values of sequence (4.19).

The kernel of integration in (4.21) represents the conditional density function of the maximum-likelihood estimator $\hat{\alpha}$ given α for fixed and finite values of the sample-size k. If the asymptotic normal density (4.10) is used as the kernel of integration instead of (4.16), then an alternative smooth estimator for α_n can be given by

$$\tilde{\alpha}_N = \frac{\displaystyle\sum_{i=1}^{n} \int_{\hat{\alpha}_{(1)}}^{\hat{\alpha}_{(n)}} e^{-\frac{k}{2}\left(\frac{\hat{\alpha}_n - \alpha}{\alpha}\right)^2} \left[\frac{\sin\left(\frac{\alpha-\hat{\alpha}_i}{2h}\right)}{\left(\frac{\alpha-\hat{\alpha}_i}{2h}\right)}\right]^2 d\alpha}{\displaystyle\sum_{i=1}^{n} \int_{\hat{\alpha}_{(1)}}^{\hat{\alpha}_{(n)}} \frac{1}{\alpha}\, e^{-\frac{k}{2}\left(\frac{\hat{\alpha}_n - \alpha}{\alpha}\right)^2} \left[\frac{\sin\left(\frac{\alpha-\hat{\alpha}_i}{2h}\right)}{\left(\frac{\alpha-\hat{\alpha}_i}{2h}\right)}\right]^2 d\alpha} . \tag{4.22}$$

For small sample sizes which are usually encountered in practical situations, the estimator $\tilde{\alpha}_N$ is shown in section 4.6 to provide inferior results when compared with $\tilde{\alpha}_G$. This is not too surprising since $\tilde{\alpha}_N$ is based on the asymptotic density of $\hat{\alpha}$ given α rather than on the true density as is $\tilde{\alpha}_G$. Therefore when β is known, we use $\tilde{\alpha}_G$ to estimate the scale parameter α. The estimator $\tilde{\alpha}_N$, however, provides substantial squared-error improvements over the corresponding maximum-likelihood estimator and for reasons to be made apparent in the following discussion has been considered here.

In some instances a closed form representation for a classical estimator $\hat{\alpha}$ will be unattainable. The asymptotic distribution of the estimator may be known to be normal, in which case $\tilde{\alpha}_N$ can be obtained. In section 4.6 the ability of $\tilde{\alpha}_N$ to provide good squared-error results is demonstrated and can be directly compared with that obtained from $\tilde{\alpha}_G$. This comparison provides some indication about the degree of departure one can expect when a smooth estimator is based on the asymptotic distribution rather than the true distribution of a classical estimator. Obviously the degree of departure depends on the rate of convergence to the asymptotic distribution as a function of k, the number of observations in each experiment. Hence no general conclusion can be reached.

4.4 Smooth Empirical Bayes Estimators for α Corrected for Variance

In section 4.3 the asymptotic distribution of the maximum-likelihood estimator $\hat{\alpha}$ given any α was shown to be

$$\text{distr.}(\hat{\alpha}|\alpha) = N\left(\alpha, \frac{\alpha^2}{k}\right) .$$

This distribution corresponds to the distribution given by (2.21). Their respective means coincide, and by setting $c = 1$ in (2.21) their variances become equivalent. Hence the transformation

$$\alpha^* = a_1\left(\hat{\alpha} - E(\hat{\alpha})\right) + E(\alpha) \tag{4.23}$$

defined by (2.25) can be performed on each element of sequence (4.19) thus forming a new sequence

$$\alpha_1^*,\ \alpha_2^*,\ \cdots,\ \alpha_n^* \tag{4.24}$$

having a marginal distribution whose mean and variance approximately equal those of the prior distribution. When the prior mean and variance of α are known, the constant a_1 in (4.23) is given by

$$a_1 = \left(\frac{k\ \mathrm{Var}(\alpha)}{(1+k)\ \mathrm{Var}(\alpha) + E^2(\alpha)}\right)^{1/2} \quad ; \tag{4.25}$$

otherwise

$$a_1 = \left[\frac{1}{1+k}\left(k - \frac{\bar{\alpha}_n^2}{s_n^2}\right)\right]^{1/2} \tag{4.26}$$

where

$$\bar{\alpha}_n = \sum_{i=1}^{n} \frac{\hat{\alpha}_i}{n} \tag{4.27}$$

and

$$s_n^2 = \frac{\sum_{i=1}^{n}\left(\hat{\alpha}_i - \bar{\alpha}_n\right)^2}{n-1} \tag{4.28}$$

Any density estimator $p_n(\alpha^*)$ formed from sequence (4.24) has the property that

$$p \lim_{\substack{n\to\infty \\ k\to\infty}} p_n(\alpha^*) = g(\alpha) \quad .$$

Therefore, when considered as a function of α , $p_n(\alpha^*)$ can be used to approximate the prior density function, and a smooth estimator for α_n becomes

$$\tilde{\alpha}_{N,V} = \frac{\displaystyle\sum_{i=1}^{n} \int_{\alpha^*_{(1)}}^{\alpha^*_{(n)}} e^{-\frac{k}{2}\left(\frac{\hat{\alpha}_n - \alpha}{\alpha}\right)^2} \left[\frac{\sin\left(\frac{\alpha-\alpha^*_i}{2h}\right)}{\left(\frac{\alpha-\alpha^*_i}{2h}\right)}\right]^2 d\alpha}{\displaystyle\sum_{i=1}^{n} \int_{\alpha^*_{(1)}}^{\alpha^*_{(n)}} \frac{1}{\alpha}\, e^{-\frac{k}{2}\left(\frac{\hat{\alpha}_n - \alpha}{\alpha}\right)^2} \left[\frac{\sin\left(\frac{\alpha-\alpha^*_i}{2h}\right)}{\left(\frac{\alpha-\alpha^*_i}{2h}\right)}\right]^2 d\alpha} \tag{4.29}$$

where $\alpha^*_{(1)}$ and $\alpha^*_{(n)}$ are the respective minimum and maximum values of sequence (4.24). Since the asymptotic normal density is used as the kernel of integration in (4.29), $\tilde{\alpha}_{N,V}$ corresponds to the smooth estimator $\tilde{\alpha}_N$ given by (4.22).

The sequence (4.24) is obtained under the assumption that the conditional distribution of the maximum-likelihood estimator is normal. For finite values of k, the actual distribution of this variable is known to be the inverted gamma distribution. Based on this distribution, a transformation similar to (4.23) can be obtained and used to develop a smooth estimator $\tilde{\alpha}_{G,V}$ which corresponds to $\tilde{\alpha}_G$. This transformation is found by following a development similar to the one used in section 2.5 for Θ^*.

The mean of the inverted gamma distribution is given by (4.17) and its variance by (4.18). If these values and the relations of conditional probability given by (2.22) and (2.23) are used, the mean and the variance for the marginal distribution of the maximum-likelihood estimator $\hat{\alpha}$ become

$$E(\hat{\alpha}) = E\left(\frac{\alpha k}{k-1}\right)$$

$$= \frac{k}{k-1} E(\alpha) \quad , \qquad (4.30)$$

and

$$\mathrm{Var}(\hat{\alpha}) = E\left[\frac{(\alpha k)^2}{(k-1)^2(k-2)}\right] + \mathrm{Var}\left(\frac{\alpha k}{k-1}\right)$$

$$= \frac{k^2}{(k-1)^2(k-2)} E(\alpha^2) + \frac{k^2}{(k-1)^2} \mathrm{Var}(\alpha)$$

$$= \frac{k^2}{(k-1)^2(k-2)}\left(\mathrm{Var}(\alpha) + E^2(\alpha)\right) + \frac{k^2}{(k-1)^2} \mathrm{Var}(\alpha)$$

$$= \frac{k^2}{(k-1)^2} \mathrm{Var}(\alpha)\left(1 + \frac{1}{k-2}\right) + \frac{k^2}{(k-1)^2(k-2)} E^2(\alpha) \tag{4.31}$$

respectively.

In order to obtain a sequence of values

$$\alpha_1', \alpha_2', \cdots, \alpha_n' \tag{4.32}$$

such that $E(\alpha') = E(\alpha)$ and $\mathrm{Var}(\alpha') = \mathrm{Var}(\alpha)$, consider the transformation of $\hat{\alpha}$ given by

$$\alpha' = c_1\left(\hat{\alpha} - E(\hat{\alpha})\right) + E(\alpha) \tag{4.33}$$

where c_1 is a constant to be determined. Now

$$E(\alpha') = c_1 E(\hat{\alpha}) - c_1 E(\hat{\alpha}) + E(\alpha) = E(\alpha) \tag{4.34}$$

and

$$\operatorname{Var}(\alpha') = c_1^2 \operatorname{Var}(\hat{\alpha}) \quad . \tag{4.35}$$

If (4.31) is substituted into (4.35) and $E(\alpha)$ is replaced by

$$E(\alpha) = \frac{k-1}{k} E(\hat{\alpha}) \quad ,$$

then the variance of α' becomes

$$\operatorname{Var}(\alpha') = c_1^2 \left[\frac{k^2}{(k-1)^2} \operatorname{Var}(\alpha) \left(1 + \frac{1}{k-2}\right) + \frac{E(\hat{\alpha})}{k-2} \right]$$

$$= c_1^2 \left[\frac{k^2 \operatorname{Var}(\alpha) + E^2(\alpha)(k-1)}{(k-1)(k-2)} \right] \tag{4.36}$$

Hence by defining

$$c_1 = \left[\frac{(k-1)(k-2) \operatorname{Var}(\alpha)}{k^2 \operatorname{Var}(\alpha) + E^2(\alpha)(k-1)} \right]^{1/2} \tag{4.37}$$

we obtain

$$\operatorname{Var}(\alpha') = \operatorname{Var}(\alpha) \quad ,$$

the desired result.

The mean and variance of the prior distribution will generally remain unknown. Thus the constant c_1 given by (4.37) cannot be exactly determined. As in section 2.5 this problem can be easily resolved by finding estimates for both $E(\alpha)$ and $\mathrm{Var}(\alpha)$. Since $E(\alpha) = \left[(k-1)/k\right]E(\hat{\alpha})$, the sample mean $\bar{\alpha}_n$ can be used to approximate $E(\hat{\alpha})$; hence $E(\alpha)$ can be estimated by

$$\widehat{E(\alpha)} = \frac{k-1}{k}\,\bar{\alpha}_n \quad . \tag{4.38}$$

If in (4.31) $\mathrm{Var}(\hat{\alpha})$ is replaced by the sample variance s_n^2 and $E(\alpha)$ is replaced by $\widehat{E(\alpha)}$, then

$$s_n^2 = \frac{k^2}{(k-1)^2}\,\mathrm{Var}(\alpha)\left(1 + \frac{1}{k-2}\right) + \frac{\bar{\alpha}_n^2}{k-2} \quad . \tag{4.39}$$

Solving (4.39) for $\mathrm{Var}(\alpha)$ we obtain

$$\widehat{\mathrm{Var}(\alpha)} = \frac{\left[(k-2)s_n^2 - \bar{\alpha}_n^2\right](k-1)}{k^2} \tag{4.40}$$

as an estimate of $\mathrm{Var}(\alpha)$. Proper substitution of (4.38) and (4.40) into (4.37) gives

$$c_1 = \frac{k-1}{k}\left[(k-2) - \frac{\bar{\alpha}_n^2}{s_n^2}\right] \quad . \tag{4.41}$$

Hence when the mean and variance of the prior distribution are unknown, (4.41) can be used to represent c_1 in the transformation (4.33).

With this transformation completely determined, sequence (4.32) is obtainable and a density estimator based on these estimates can be used to replace the prior density function giving the smooth estimator for α_n

$$\tilde{\alpha}_{G,V} = \frac{\displaystyle\sum_{i=1}^{n} \int_{\alpha'_{(1)}}^{\alpha'_{(n)}} \left(\frac{\alpha k}{\hat{\alpha}_n}\right)^{k+1} e^{-\left(\frac{\alpha k}{\hat{\alpha}_n}\right)} \left[\frac{\sin\left(\frac{\alpha-\alpha'_i}{2h}\right)}{\left(\frac{\alpha-\alpha'_i}{2h}\right)}\right]^2 d\alpha}{\displaystyle\sum_{i=1}^{n} \int_{\alpha'_{(1)}}^{\alpha'_{(n)}} \frac{1}{\alpha}\left(\frac{\alpha k}{\hat{\alpha}_n}\right)^{k+1} e^{-\left(\frac{\alpha k}{\hat{\alpha}_n}\right)} \left[\frac{\sin\left(\frac{\alpha-\alpha'_i}{2h}\right)}{\left(\frac{\alpha-\alpha'_i}{2h}\right)}\right]^2 d\alpha} \tag{4.42}$$

The limits of integration $\alpha'_{(1)}$ and $\alpha'_{(n)}$ are the respective minimum and maximum of sequence (4.32).

4.5 Iteration of the Smooth Estimators

Consider a sequence of smooth estimates from (4.21)

$$\tilde{\alpha}_{G,1}, \tilde{\alpha}_{G,2}, \cdots, \tilde{\alpha}_{G,n} \tag{4.43}$$

obtained in each of n past experiments. Based on this sequence, a more "precise" estimate of the realization α_n can often be determined. This estimate is obtained by replacing the prior density in the Bayes estimator by a density approximation constructed with sequence (4.43). This approach represents an iteration of the smooth estimator $\tilde{\alpha}_G$ and is denoted by $\tilde{\alpha}'_G$ since the kernel of integration remains unchanged. Similar iterations of the estimators $\tilde{\alpha}_{G,V}$, $\tilde{\alpha}_N$, and $\tilde{\alpha}_{N,V}$ can be obtained and will be denoted by $\tilde{\alpha}'_{G,V}$, $\tilde{\alpha}'_N$, and $\tilde{\alpha}'_{N,V}$ respectively. Integration in each estimator is performed over the range of values obtained from the first iteration. For example, the region of integration for $\tilde{\alpha}'_G$ is from $\min(\tilde{\alpha}_{G,i})$ to $\max(\tilde{\alpha}_{G,j})$ where $i,j = 1,2,\cdots,n$ and $i \neq j$. For $i = 1$, we define $\tilde{\alpha}_{G,1} = \hat{\alpha}_1$.

4.6 Monte Carlo Simulation

The continuously smooth empirical Bayes estimators given in the previous sections were investigated for small samples by Monte Carlo simulation and compared to maximum-likelihood estimators. The simulation was conducted in a manner analogous to the procedure described in section 3.5. In each experiment a value of α was generated from a prior distribution belonging to the Pearson family of distributions. This parameter was then used to obtain a sample of size k from (4.1), and the maximum-likelihood and continuously smooth empirical Bayes estimates were calculated. Five hundred repetitions of each of twenty experiments were made for each prior distribution, and the ratio

$$R = \frac{\text{empirical Bayes mean-squared error}}{\text{maximum-likelihood mean-squared error}}$$

was formed. As in section 3.5, it was found that the ratio R was significantly influenced by the prior distribution only through the value

$$Z = \frac{\text{Var}^*(\hat{\alpha}|\alpha)}{\text{Var}(\alpha)}$$

where * indicates that $E(\alpha)$, the prior mean of α, has been substituted for α. In particular for a random sample of size k from (4.1), $\text{Var}(\hat{\alpha}|\alpha)$ is given by (4.18), and the value Z becomes

$$Z = \frac{k^2 E^2(\alpha)}{(k-1)^2 (k-2)\ \text{Var}(\alpha)} . \tag{4.44}$$

Since the only factors affecting the ratio R, apart from the number of experiences, are contained in (4.44), this quantity can be conveniently used to summarize and index a given situation.

As in Chapter III we wish to illustrate the robustness of the continuously smooth empirical Bayes estimators to the form of the prior distribution. Accordingly, values of $\alpha_i (i = 1,2,\cdots,20)$ were generated from various Pearson prior distributions. The ratio R was observed to vary only slightly for a given value of n, providing the value of Z remained invariant from distribution to distribution. Illustrations of this fact are presented in Figures 12-15 for a given value of $Z = 2.0$. In Figures 12, 13, 14, and 15 the prior distribution is bell-shaped (skewed), L-shaped, J-shaped, and U-shaped respectively. The solid line in each figure represents the ratio $R_{G,V}$ calculated with the smooth estimator $\tilde{\alpha}_{G,V}$ given by (4.42); the broken line represents the ratio R_G calculated with the smooth estimator $\tilde{\alpha}_G$ given by (4.21). This representation is used in the remaining figures of this chapter. The value Z is the same for all four figures, but skewness (S) and kurtosis (K) vary, giving different forms to the prior distributions. Little difference, however, can be detected in the corresponding values of $R_{G,V}$ and R_G in the four figures. Hence the smooth empirical Bayes estimators are quite insensitive to the form of the prior distribution as summarized by S and K. Therefore the values of S and K will not be given for the remaining figures in this chapter. The values of the prior distribution are designated as follows:

$$E(\alpha) \;=\; \text{prior mean of } \alpha \;;$$

$$V(\alpha) \;=\; \text{prior variance of } \alpha \;.$$

To determine the effect of the sample size k on the ratios $R_{G,V}$ and R_G, several runs were made with k ranging from 5 to 20. Results from these investigations revealed that the quantities in Z can vary in a manner not affecting the value of Z without having a significantly noticeable effect on the ratios $R_{G,V}$ and R_G. That is, the value of Z and not the individual quantities in Z determines the values of $R_{G,V}$ and R_G. Figures 16, 17, and 18 illustrate this point. By comparing these figures, it can be seen that although the parameters k, $E(\alpha)$, and $V(\alpha)$ vary widely in each figure, with the value of Z remaining 3.5, the ratios $R_{G,V}$ and R_G remain relatively unchanged. Since the ratio R remains relatively unchanged for equivalent values of Z, the individual quantities in Z will not be given for the remaining figures in this chapter.

Values of the ratios R_G and $R_{G,V}$ are plotted in Figures 19 and 20 respectively. These ratios are plotted for different values of Z ranging from 0.5 to 5.0. Again we notice that as Z increases, the ratios $R_{G,V}$ and R_G decrease. In particular for a given value of Z, the values of $R_{G,V}$ are smaller than those of R_G, demonstrating the superiority of the smooth estimator $\tilde{\alpha}_{G,V}$ over the smooth estimator $\tilde{\alpha}_G$. Again this increase in mean-squared precision is attributed to the prior density approximation used in $\tilde{\alpha}_{G,V}$. This approximation is based on a sequence of values having a marginal distribution whose mean and variance are approximately equivalent to those of the prior distribution.

Figure 21 represents a typical result obtained when $\tilde{\alpha}_{G,V}$ is iterated as described in section 4.5. The ratio $R'_{G,V}$ formed with the iterated smooth estimator $\tilde{\alpha}'_{G,V}$ is represented by the dotted line. As in Chapter III we notice that the squared-error improvement achieved by using $\tilde{\alpha}_{G,V}$ is slightly decreased by a second iteration. This decrease in precision is expected. The approximation used to represent the prior density in $\tilde{\alpha}_{G,V}$ is based on a sequence of values whose marginal distribution has its mean and variance approximately equivalent to those of the prior distribution. The prior density approximation in $\tilde{\alpha}'_{G,V}$, however, is not known to have this property.

Results from a second iteration of $\tilde{\alpha}_G$ are given in Figure 22 for a value of Z identical to that used in Figure 21. The ratio R_G' formed with the smooth estimator $\tilde{\alpha}_G'$ is represented by the dotted line. We note that iteration of $\tilde{\alpha}_G$ increases the mean-squared precision; however, this improvement is not as significant as that obtained with $\tilde{\alpha}_{G,V}$. It is conjectured that the over-estimation of the prior variance by the marginal distribution of the maximum-likelihood estimator has a far more significant effect on the prior density approximation than does the marginal variance of $\tilde{\alpha}_G$.

In the Monte Carlo study the estimator $\tilde{\alpha}_{G,V}$ provided uniform squared-error improvement over the maximum-likelihood estimator for two or more experiences. It was also observed to be the most efficient of the smooth estimators. Since this improvement was consistent over a wide variety of Z values, we recommend that the smooth estimator $\tilde{\alpha}_{G,V}$ be used in all situations.

In Figure 23 the ratio $R_{G,V}$ is plotted as a function of Z for a given number of experiences. These plots give some indication of the improvement over maximum-likelihood one can expect if an estimate for the value of Z can be obtained.

As discussed in section 4.3, in some instances the true density function of the maximum-likelihood estimator for a finite sample size k may be unknown. Its asymptotic distribution may, however, be known and can be used to form an efficient smooth estimator. The estimators $\tilde{\alpha}_N$ and $\tilde{\alpha}_{N,V}$ given by (4.22) and (4.29) respectively, were constructed in this manner. Values of the ratios R_N formed with $\tilde{\alpha}_N$ and $R_{N,V}$ formed with $\tilde{\alpha}_{N,V}$ are plotted in Figures 24 and 25 respectively. These ratios are plotted for various values of Z ranging from 0.5 to 5.0. Comparison of Figures 24 and 25 with Figure 20 shows that the estimators $\tilde{\alpha}_N$ and $\tilde{\alpha}_{N,V}$ are not as "good" as the estimator $\tilde{\alpha}_{G,V}$. However they do provide significant and uniform squared-error improvement over the maximum-likelihood estimator for two or more experiences, and this is of paramount importance.

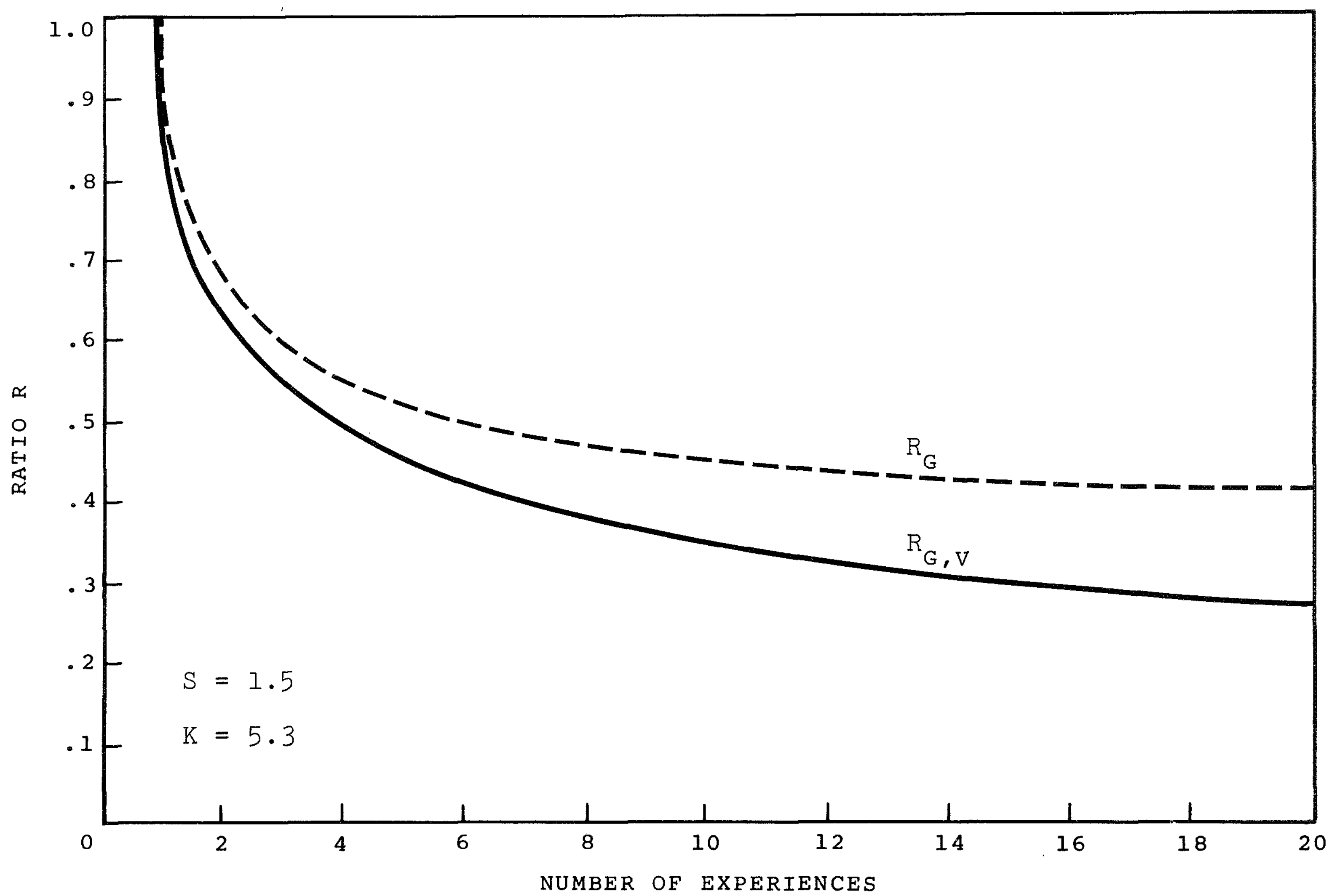

Figure 12. — R_G and $R_{G,V}$ for $Z = 2.0$; bell-shaped (skewed) prior.

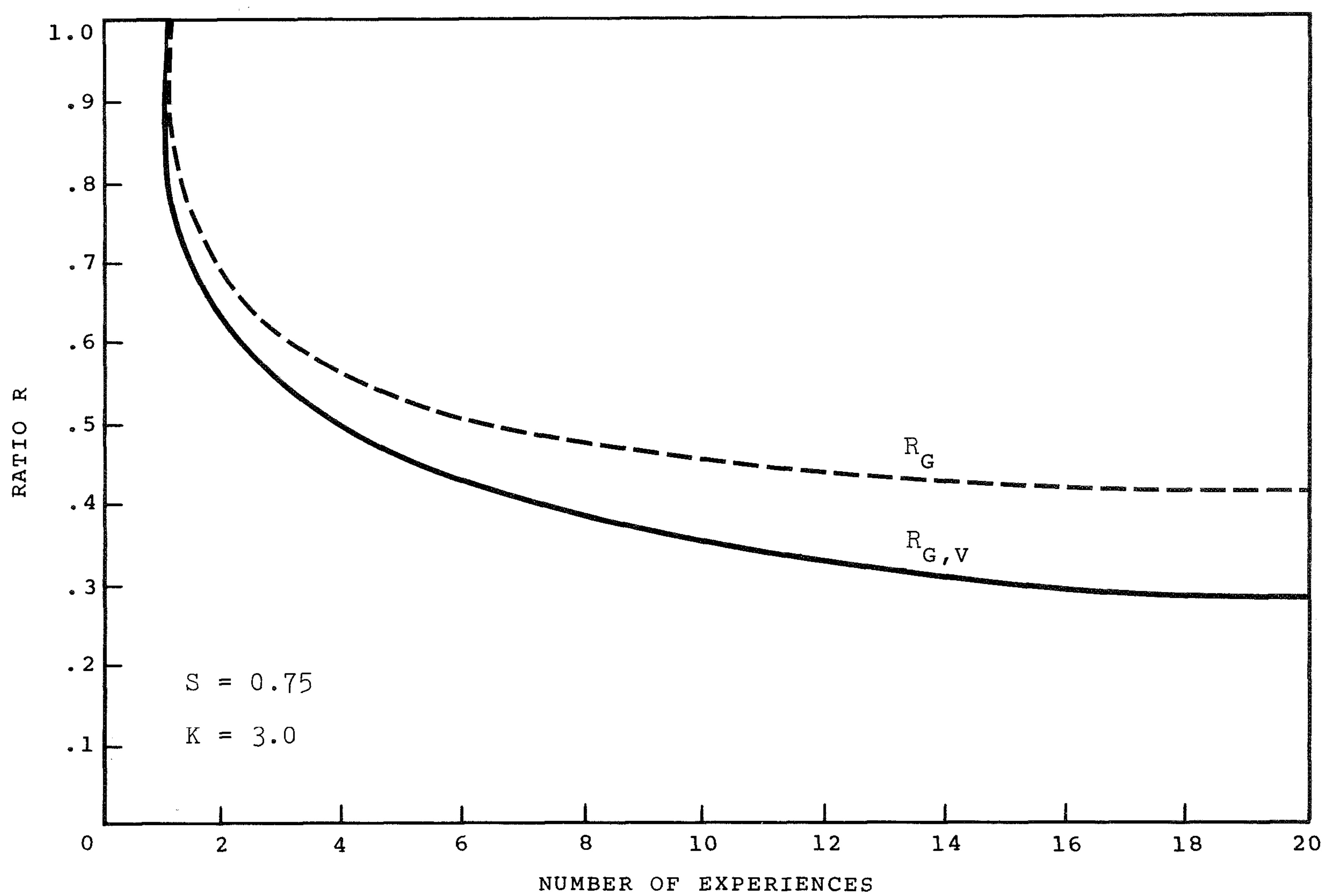

Figure 13. — R_G and $R_{G,V}$ for Z - 2.0 ; L-shaped prior.

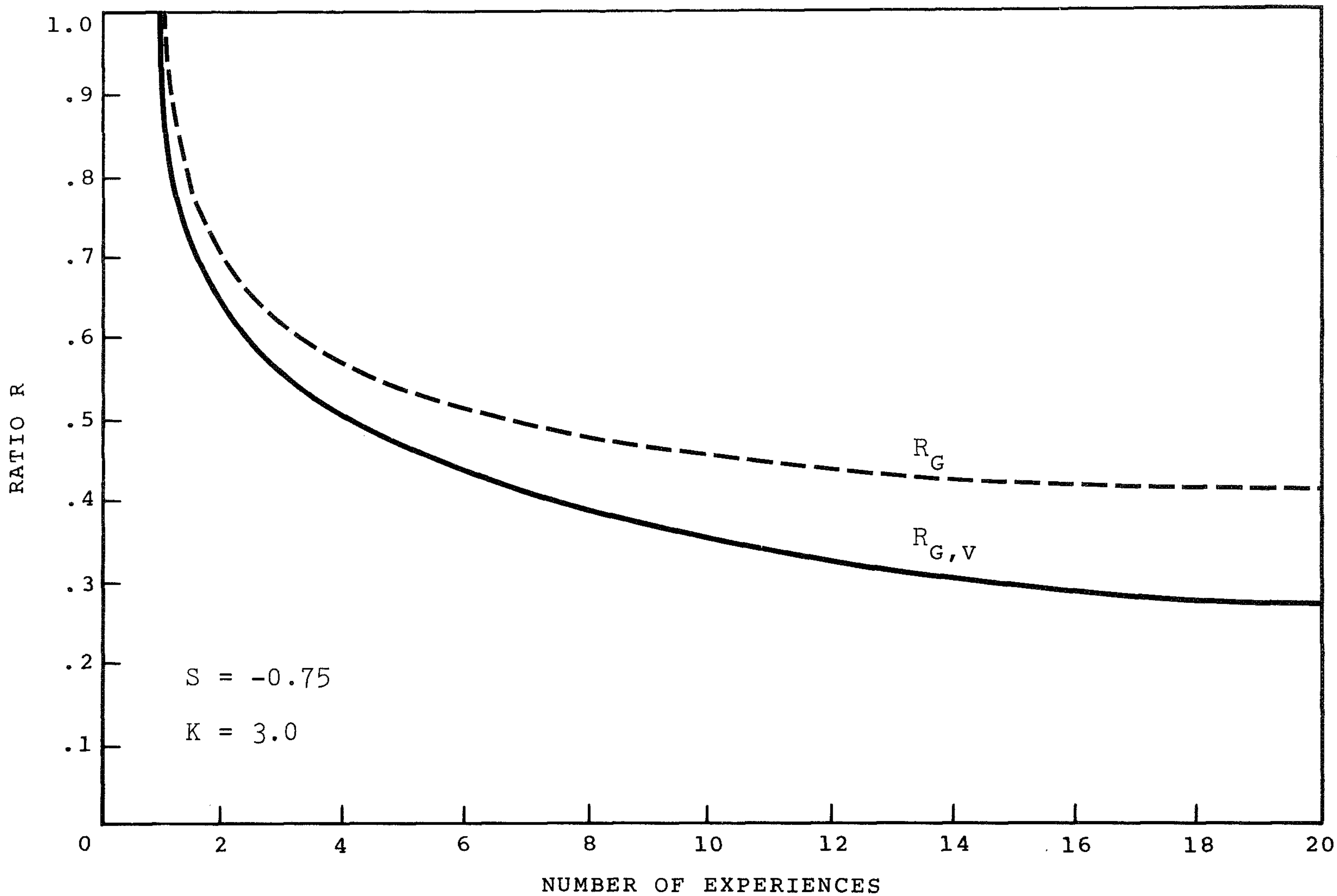

Figure 14. – R_G and $R_{G,V}$ for $Z = 2.0$; J-shaped prior.

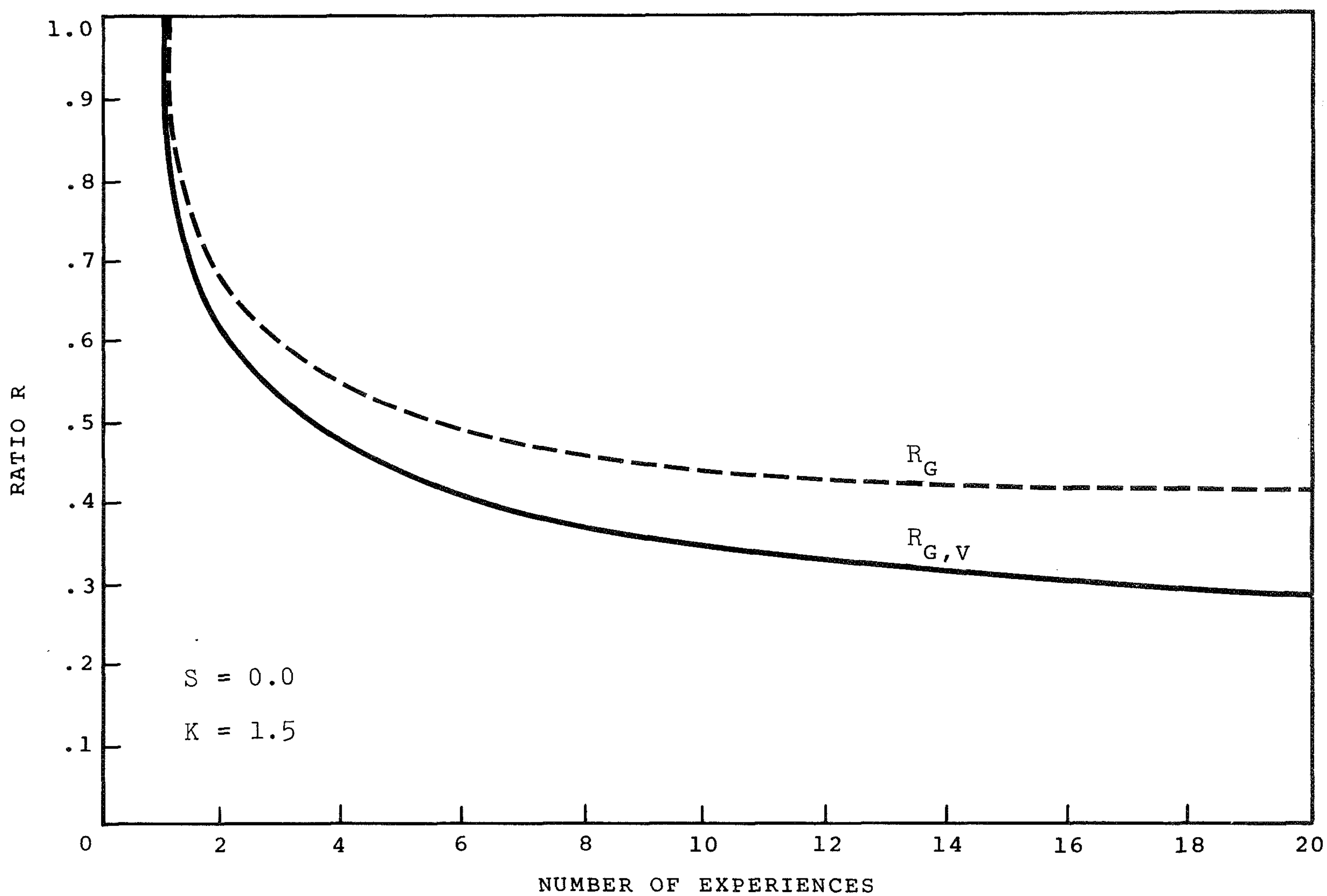

Figure 15. — R_G and $R_{G,V}$ for $Z = 2.0$; U-shaped prior.

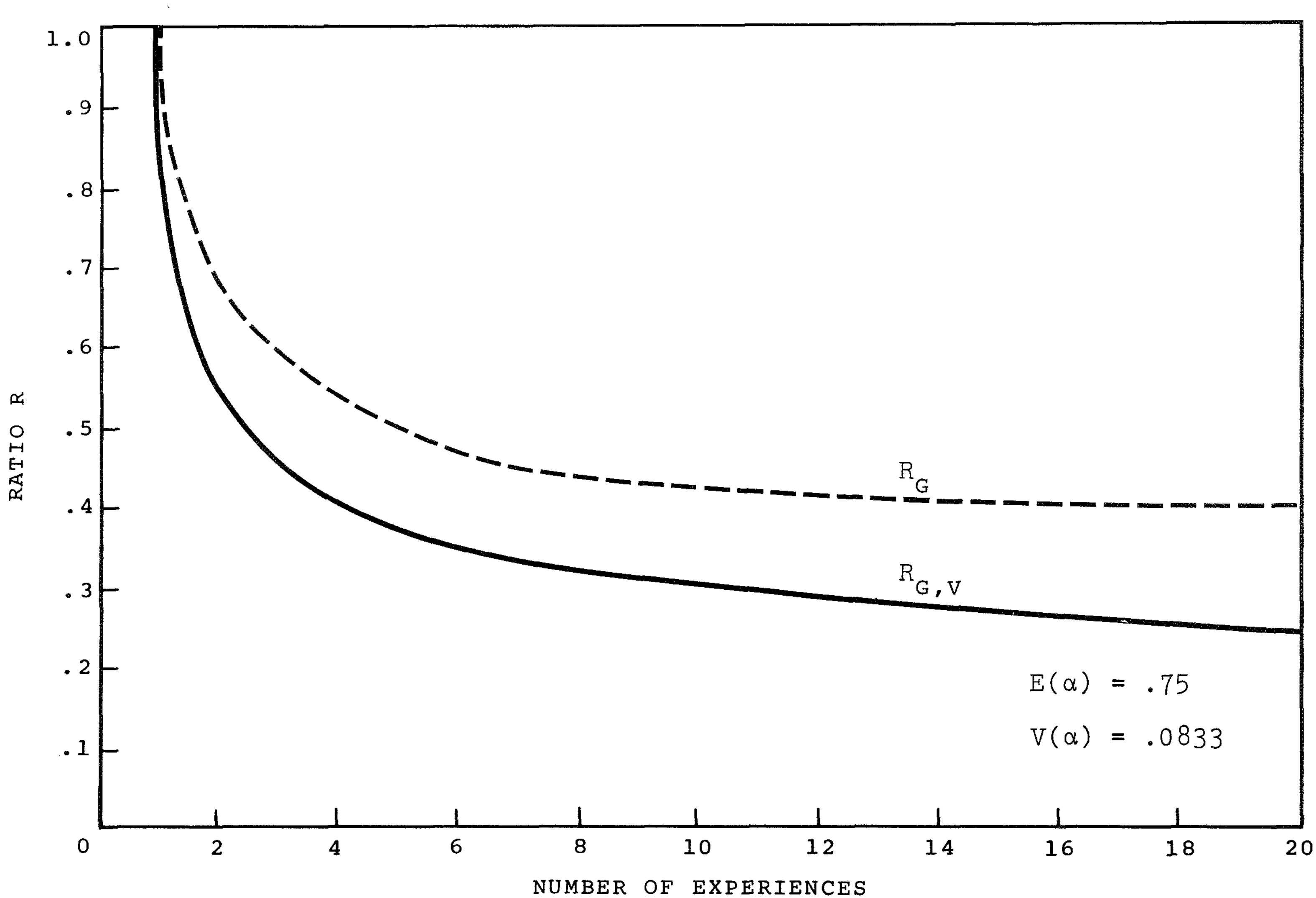

Figure 16. — R_G and $R_{G,V}$ for $Z = 3.52$ with 5 observations.

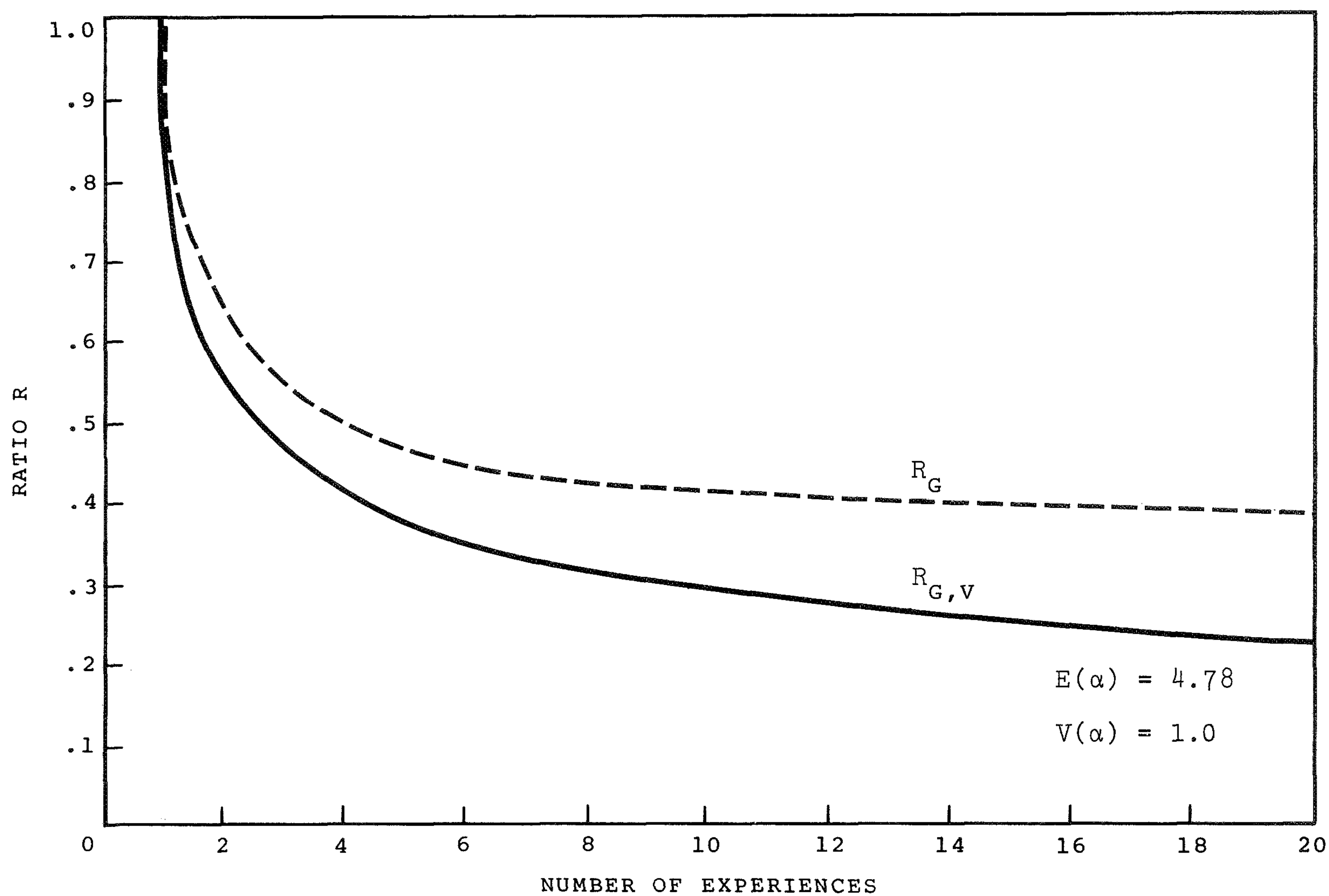

Figure 17. – R_G and $R_{G,V}$ for $Z = 3.52$ with 10 observations.

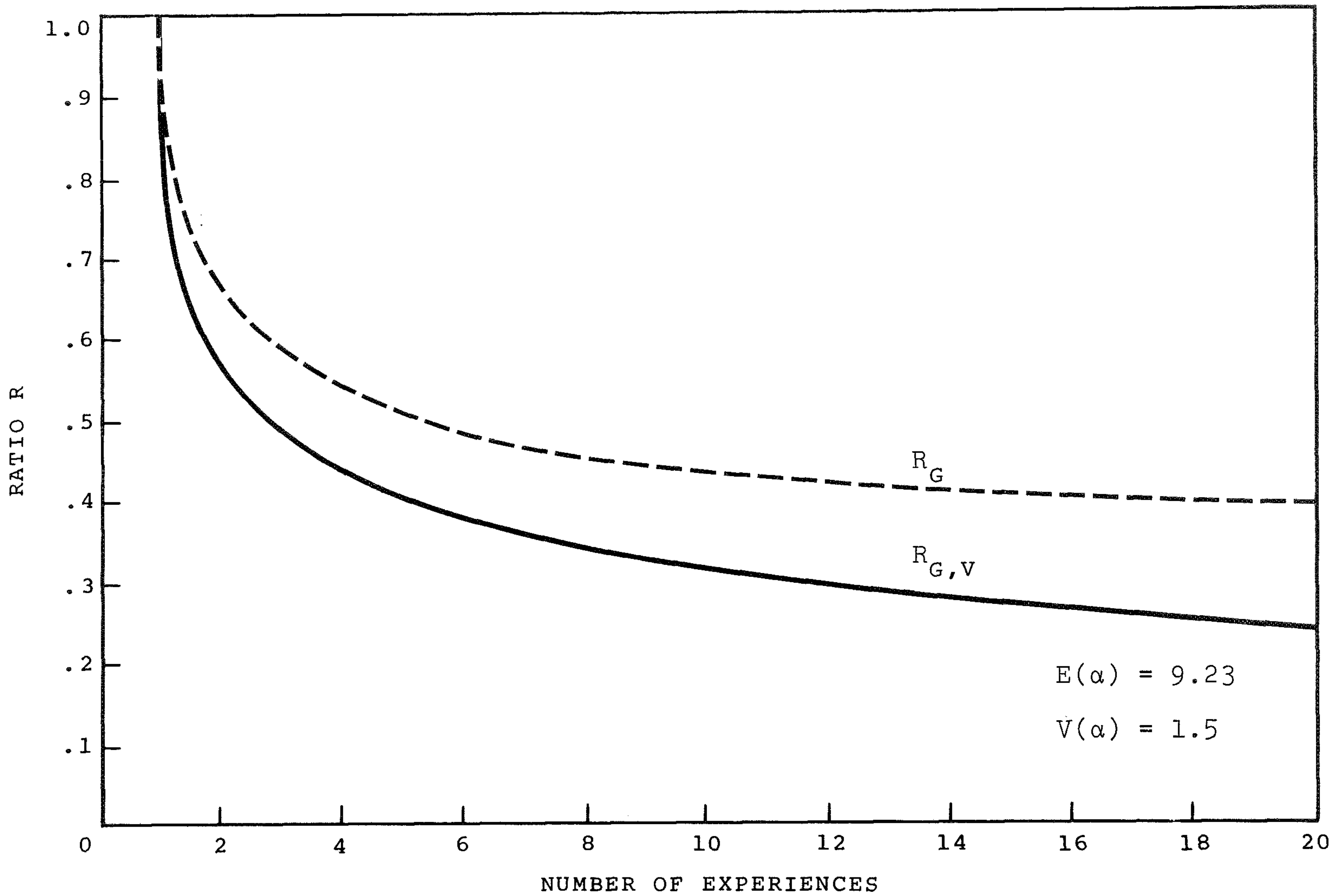

Figure 18. — R_G and $R_{G,V}$ for $Z = 3.52$ with 20 observations.

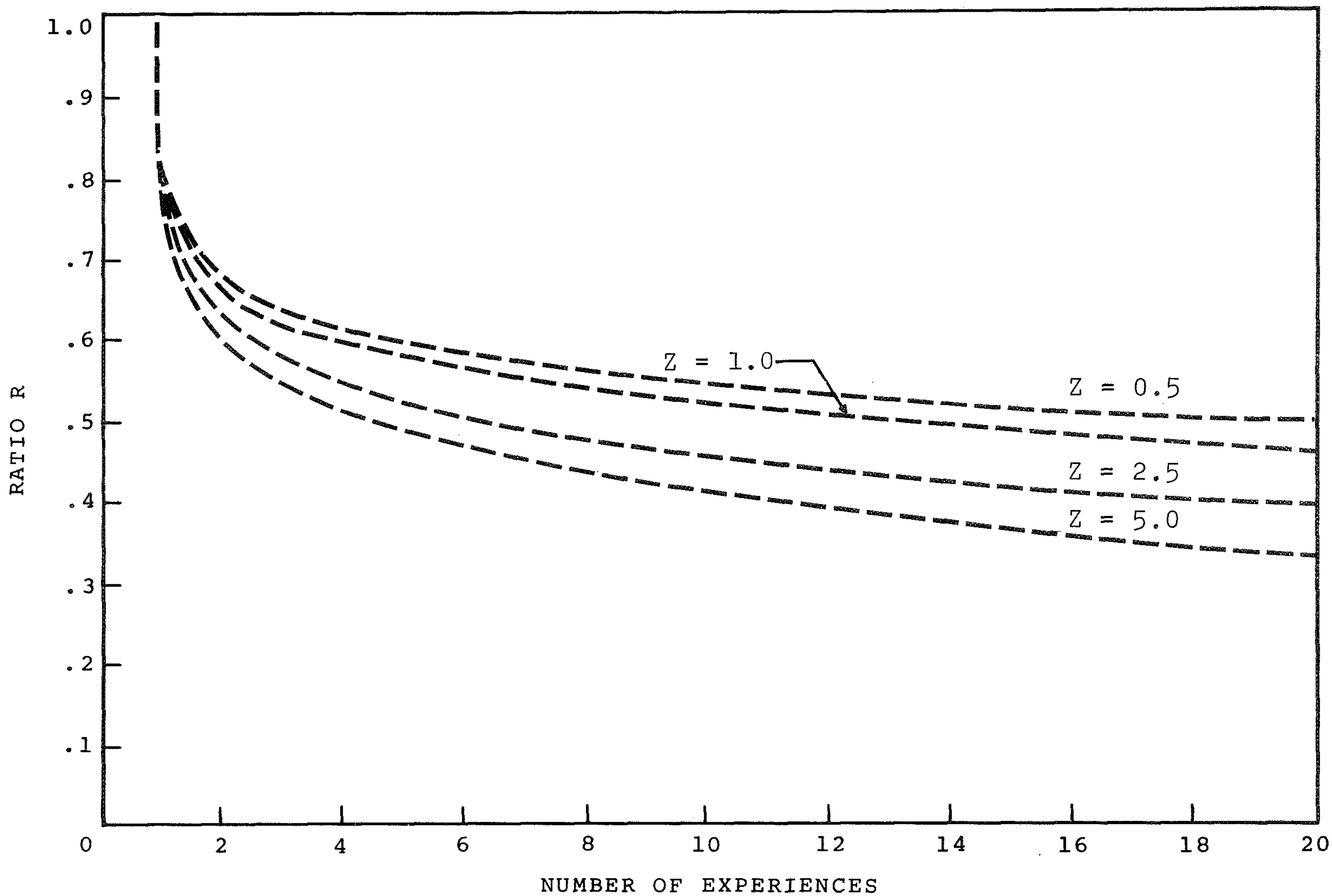

Figure 19. — R_G for several values of Z .

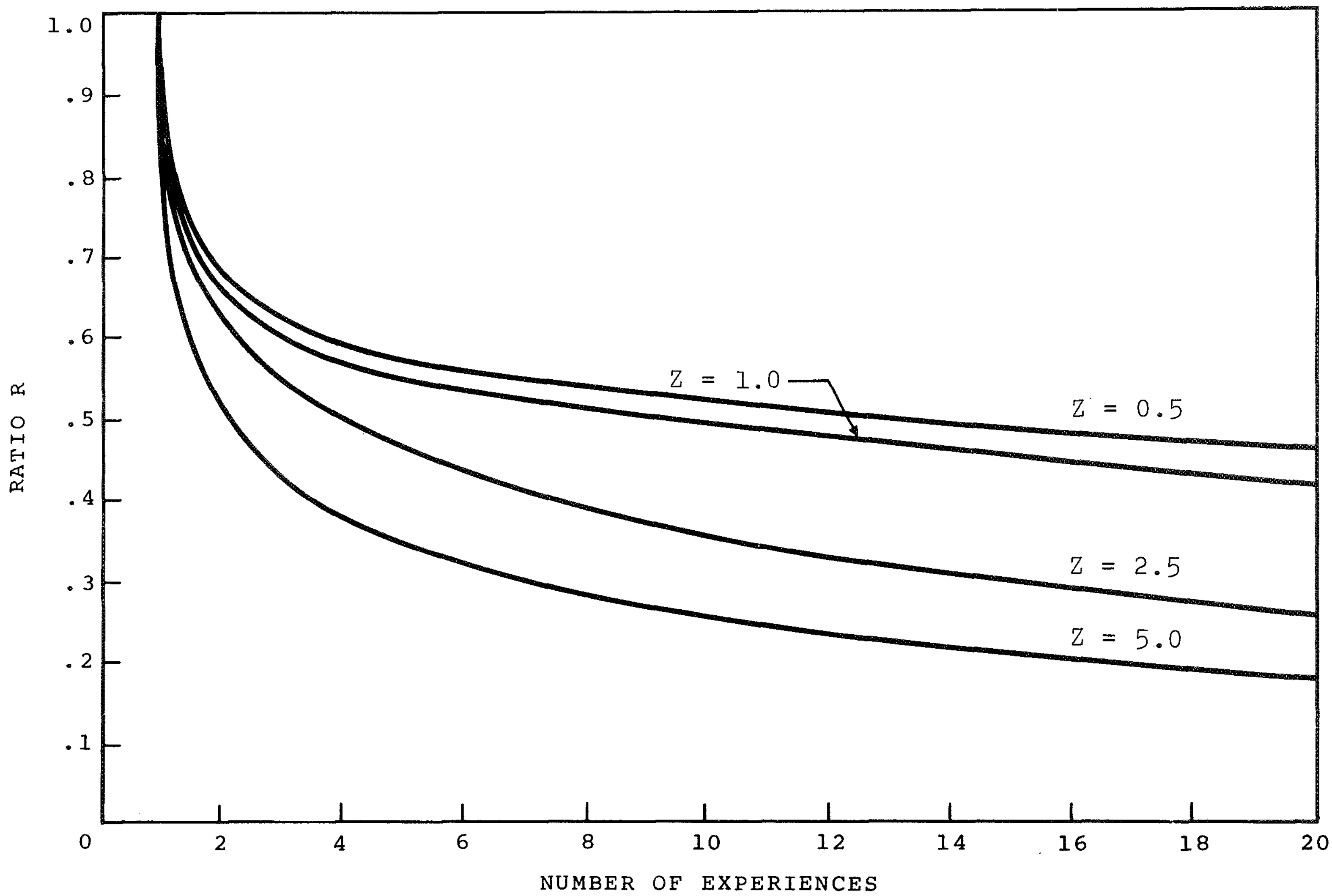

Figure 20. — $R_{G,V}$ for several values of Z.

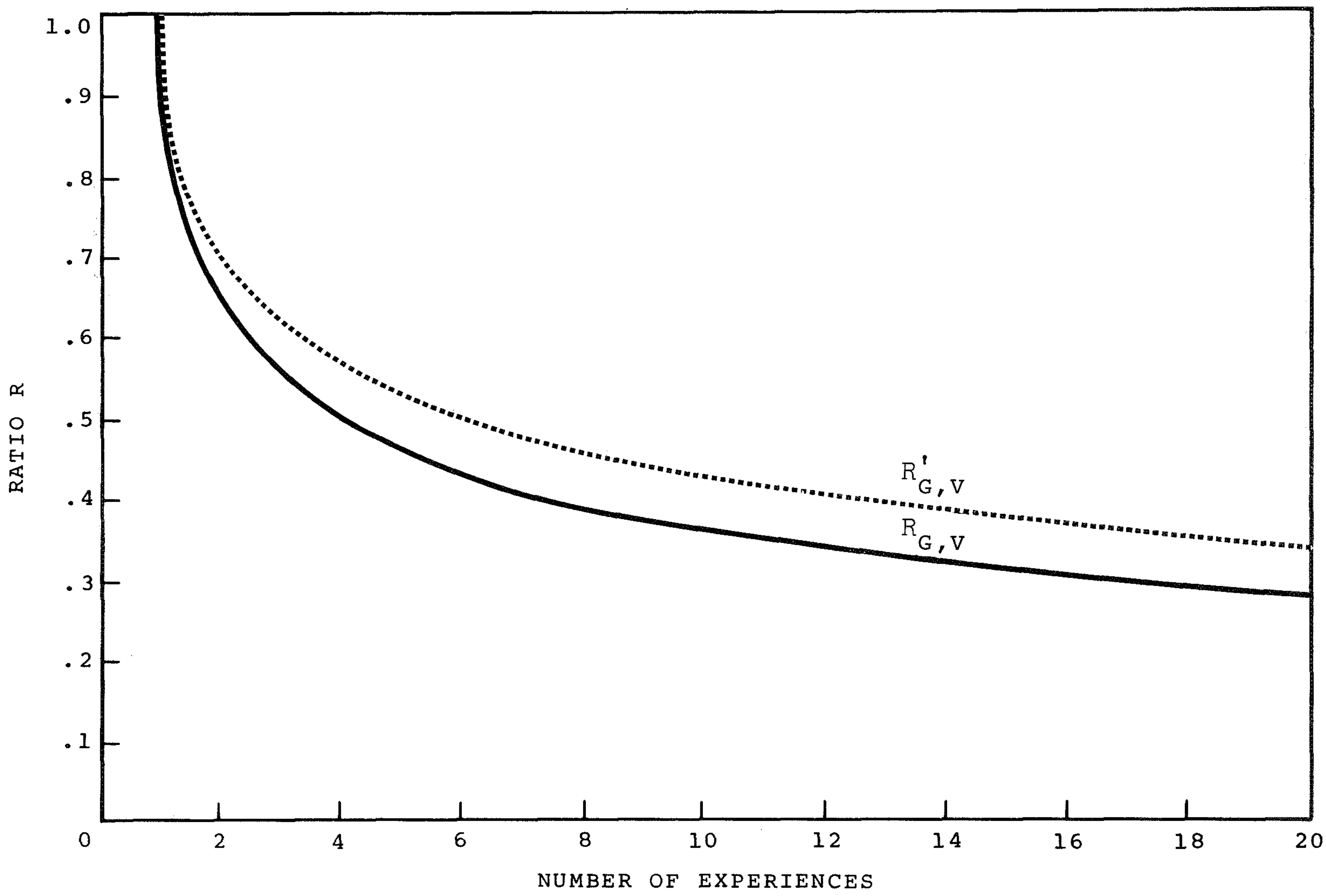

Figure 21. — $R_{G,V}$ vs $R'_{G,V}$ for $Z = 2.0$.

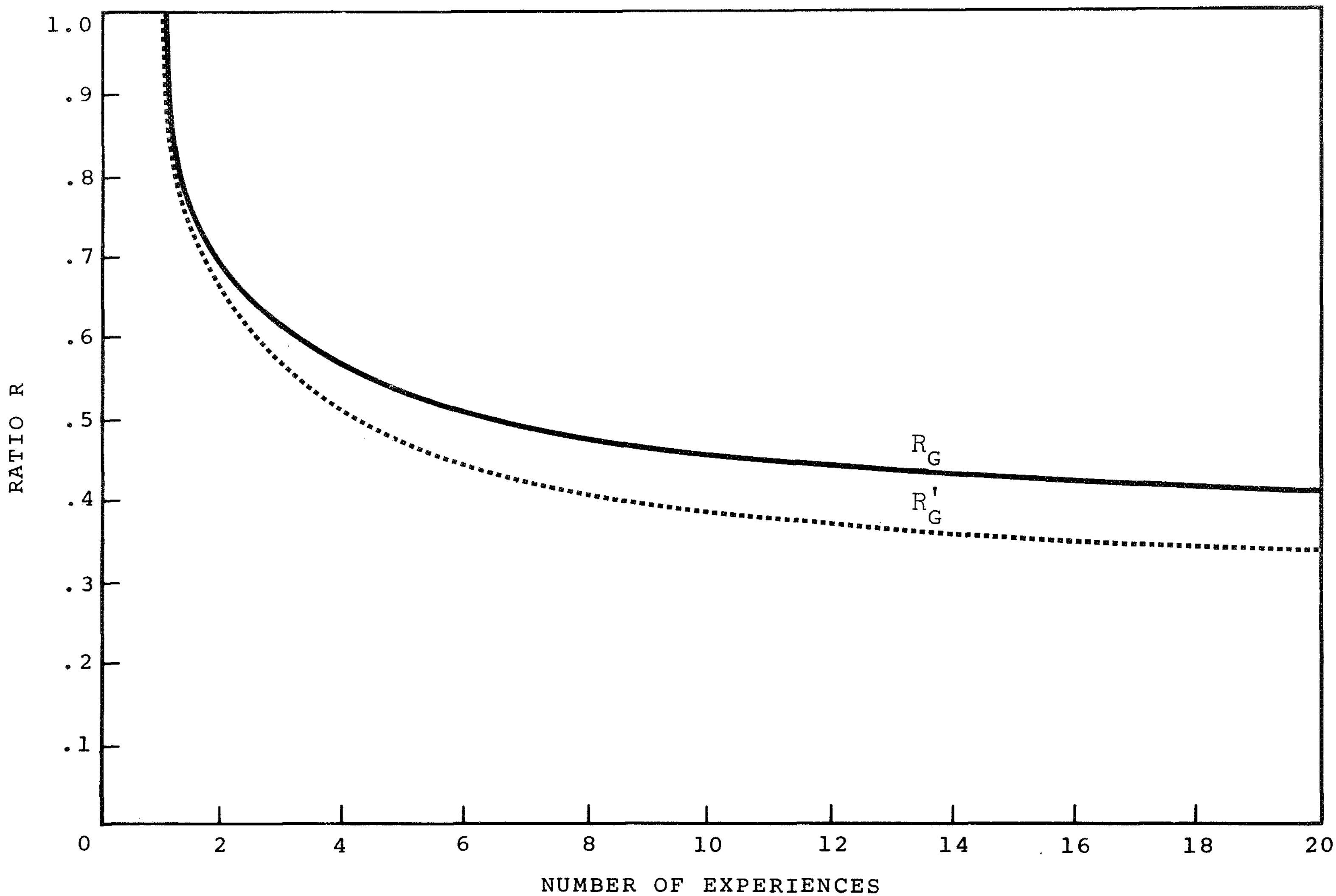

Figure 22. — R_G vs R'_G for $Z = 2.0$.

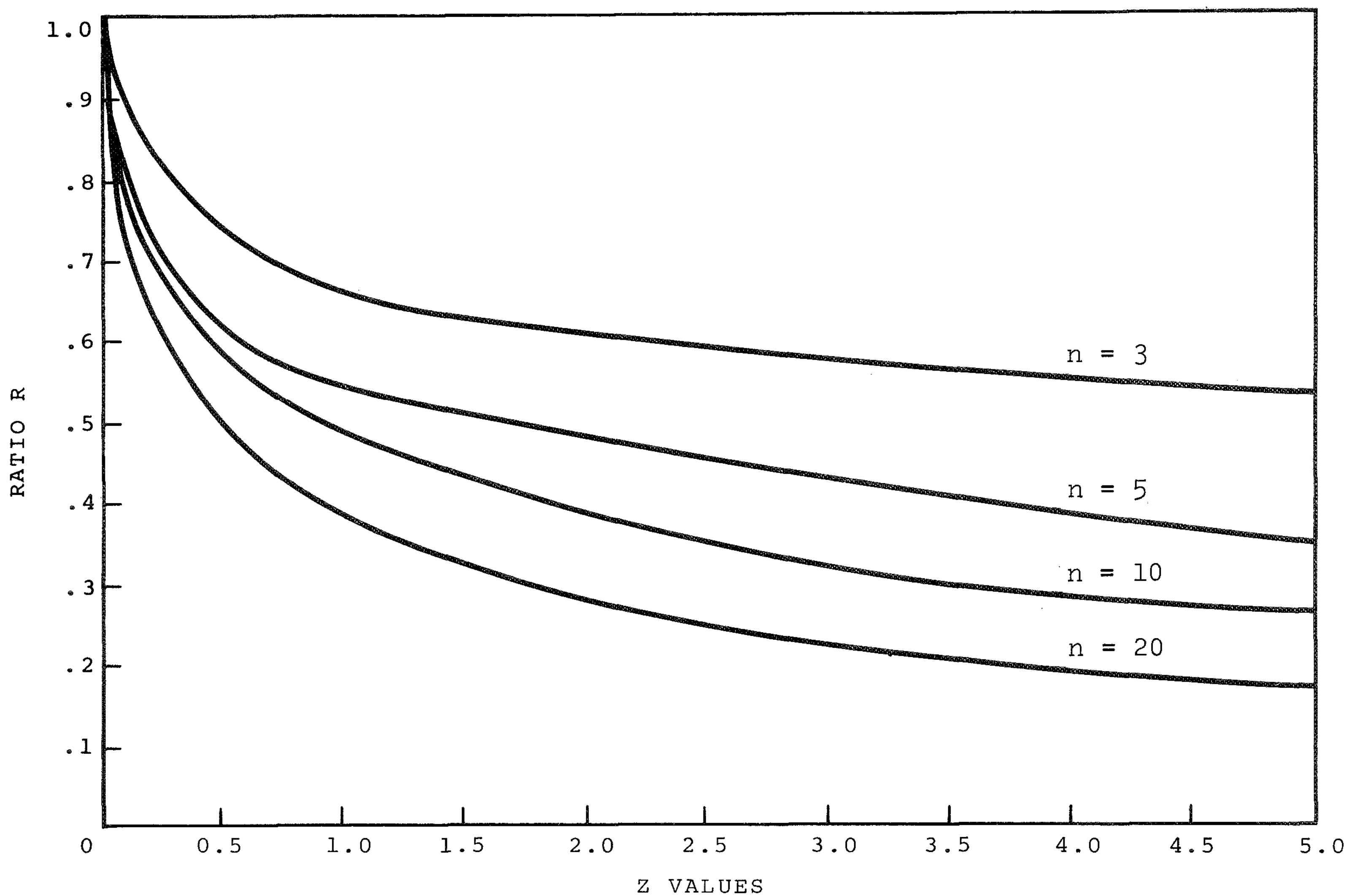

Figure 23. — $R_{G,V}$ as a function of Z for fixed number of experiences.

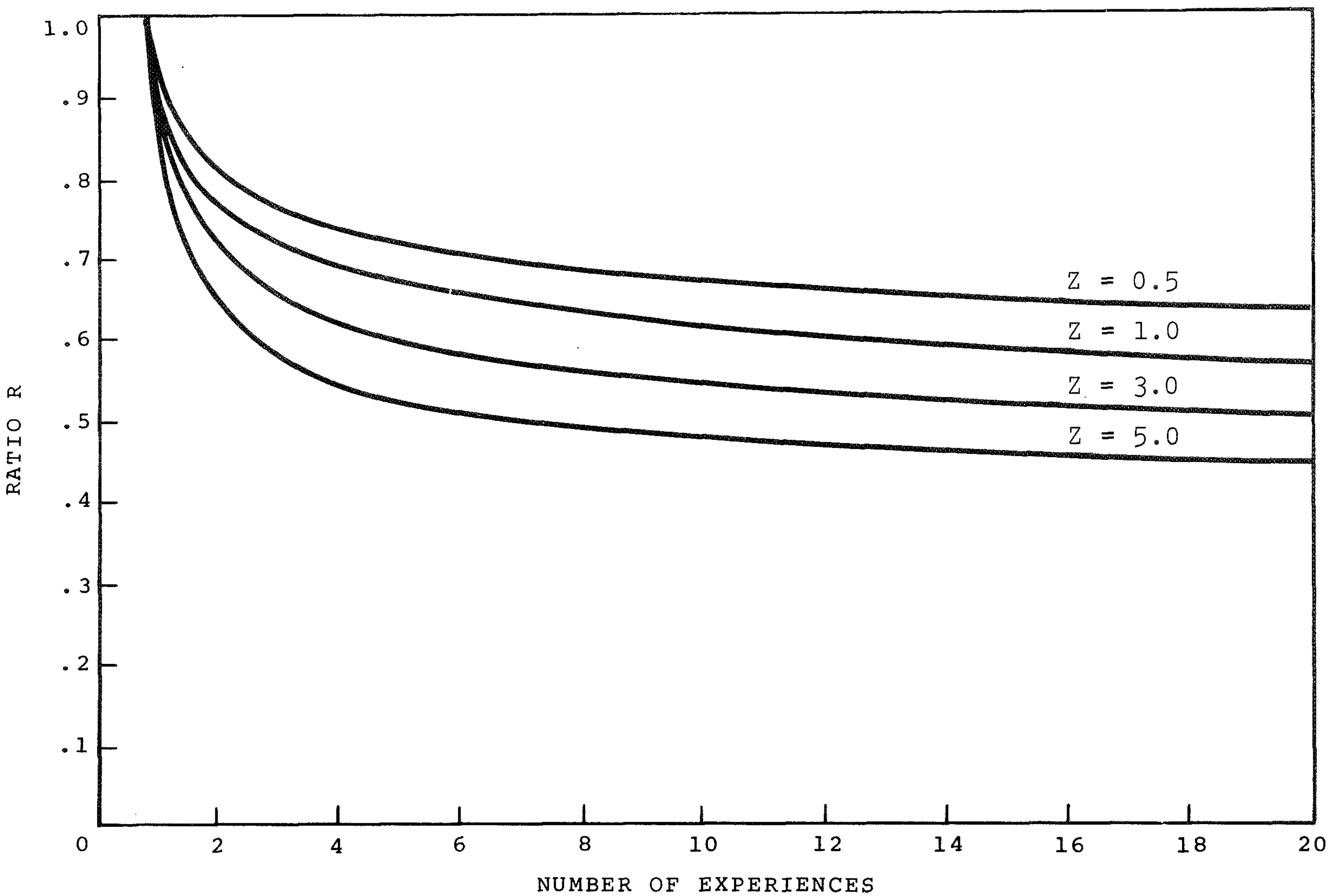

Figure 24. — R_N for several values of Z .

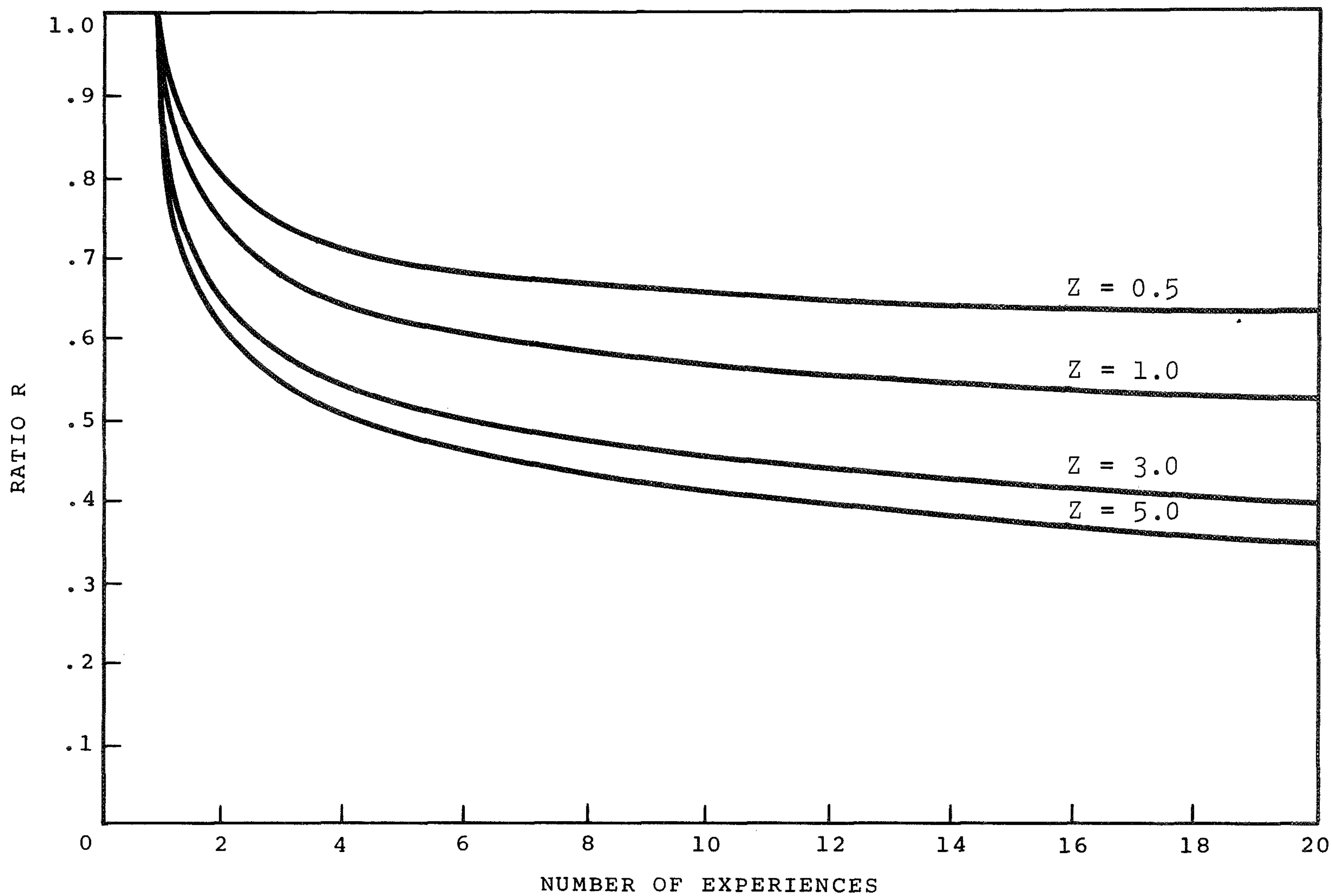

Figure 25. – $R_{N,V}$ for several values of Z .

CHAPTER V

ESTIMATION OF THE WEIBULL SHAPE PARAMETER WITH KNOWN SCALE PARAMETER

In this chapter, smooth empirical Bayes estimators are given for the shape parameter β in the two-parameter Weibull distribution. The scale parameter α is assumed to be known and fixed in each experiment. Results from Monte Carlo simulations are reported which show that even for small sample sizes and few experiments the smooth estimators have smaller mean-squared errors than the maximum-likelihood estimators.

5.1 Maximum-Likelihood Estimator for β

Let X be a random variable having a Weibull distribution with a known scale parameter α. If the shape parameter β is a random variable, then the conditional density function of X is given by

$$f(x|\beta) = \alpha\beta x^{\beta-1}e^{-\alpha x^{\beta}} , \quad (x \geq 0; \alpha, \beta > 0) . \tag{5.1}$$

Consider a random sample of k observations from (5.1). The likelihood function of this sample is

$$L(\underset{\sim}{x}|\beta) = (\alpha\beta)^k \prod_{i=1}^{k} x_i^{\beta-1} e^{-\alpha x_i^\beta} \tag{5.2}$$

Upon taking the logarithm of (5.2), differentiating with respect to β, and equating to zero, we have

$$\frac{d \log L(\underset{\sim}{x}|\beta)}{d\beta} = \frac{k}{\beta} + \sum_{i=1}^{k} \log x_i - \alpha \sum_{i=1}^{k} x_i^\beta \log x_i = 0 . \tag{5.3}$$

This equation can be solved to obtain the maximum-likelihood estimator $\hat{\beta}$. This may be accomplished with the aid of standard iterative techniques. In Appendix B such a procedure is described.

The maximum-likelihood estimator is not known to be sufficient for β, and its exact distributional form is unknown. The estimator is, however, consistent for β and by virtue of (1.25) is distributed asymptotically normal, with mean β and variance given by

$$\operatorname{Var}(\hat{\beta}|\beta) = -\left[E_x\left(\frac{d^2 \log L(\underset{\sim}{x}|\beta)}{d\beta^2}\right)\right]^{-1} \tag{5.4}$$

Now

$$\frac{d^2 \log L(\underset{\sim}{x}|\beta)}{d\beta^2} = \frac{k}{\beta^2} - \alpha\sum_{i=1}^{k} x_i^{\beta}(\log x_i)^2 \qquad (5.5)$$

and the expected value of (5.5) is

$$E\left(\frac{d^2 \log L(\underset{\sim}{x}|\beta)}{d\beta^2}\right) = -\frac{k}{\beta^2} - \alpha k E\left[x^{\beta}(\log x)^2\right] . \qquad (5.6)$$

Hence (5.4) becomes

$$\mathrm{Var}(\hat{\beta}|\beta) = \left[\frac{k}{\beta^2} + \alpha k E\left[x^{\beta}(\log x)^2\right]\right]^{-1} . \qquad (5.7)$$

Consider

$$E\left[x^{\beta}(\log x)^2\right] = \int_0^{\infty} x^{\beta}(\log x)^2 \alpha\beta x^{\beta-1} e^{-\alpha x^{\beta}} \, d\beta . \qquad (5.8)$$

If the substitution $u = x^{\beta}$ is made then

$$E\left[x^{\beta}(\log x)^2\right] = \int_0^{\infty} \frac{\alpha}{\beta^2} u \log (u) e^{-\alpha u} du \qquad (5.9)$$

which after integrating by parts, becomes

$$E\left[x^{\beta}(\log x)^2\right] = \frac{1}{\beta^2}\left[2\int_0^\infty e^{-\alpha u}\log u\,du + \int_0^\infty e^{-\alpha u}\log^2 u\,du\right]$$

$$= \frac{1}{\alpha\beta^2}\left[\frac{\pi^2}{6} + \left(\log\alpha - \psi(1)\right)^2 - 2\left(\log\alpha - \psi(1)\right)\right] \quad . \tag{5.10}$$

The function $\psi(1)$ in (5.10) is the digamma function $\psi(x)$ evaluated at $x = 1$. Completing the square in (5.10) and using the recurrence formula $\psi(x+1) = 1/x + \psi(x)$, we obtain

$$E\left[x^{\beta}(\log x)^2\right] = \frac{1}{\alpha\beta^2}\left[\frac{\pi^2}{6} + \left(\log\alpha - \psi(1) - 1\right)^2 - 1\right]$$

$$= \frac{1}{\alpha\beta^2}\left[\frac{\pi^2}{6} + \left(\psi(2) - \log\alpha\right)^2 - 1\right] \quad . \tag{5.11}$$

Hence (5.7) becomes

$$\operatorname{Var}(\hat{\beta}|\beta) = \frac{\beta^2}{k}\left[\frac{\pi^2}{6} + \left(\psi(2) - \log\alpha\right)^2\right]^{-1} . \qquad (5.12)$$

5.2 Smooth Empirical Bayes Estimators for β

Assume that the Weibull distribution with known scale parameter α adequately describes data obtained in each of n previous experiments. Further assume, that between successive experiments, the shape parameter β varies randomly with the same but unknown prior density function $g(\beta)$. If in each experiment a maximum-likelihood estimate $\hat{\beta}_i$ $(i = 1,2,\cdots,n)$ was obtained for each realization from $g(\beta)$, then the sequence of estimates

$$\hat{\beta}_1,\ \hat{\beta}_2,\ \cdots,\ \hat{\beta}_n \qquad (5.13)$$

can be used to form an approximation to the marginal density. This approximation can be represented by

$$p'_n(\hat{\beta}) = \frac{1}{2\pi nh}\sum_{i=1}^{n}\left[\frac{\sin\left(\frac{\hat{\beta}-\hat{\beta}_i}{2h}\right)}{\left(\frac{\hat{\beta}-\hat{\beta}_i}{2h}\right)}\right]^2 \qquad (5.14)$$

where h is given by (2.8).

Convergence in probability of (5.14) to the prior density $g(\beta)$ as both the sample size k and the number of experiments n tend to infinity is assured by Theorem 2.2. Therefore, when the kernel of integration is normal with mean β and variance given by (5.12), $p_n(\hat{\beta})$ can be considered as a function of β, and a continuously smooth empirical Bayes estimator for β_n becomes

$$\tilde{\beta}_N = \frac{\displaystyle\sum_{i=1}^{n} \int_{\hat{\beta}_{(1)}}^{\hat{\beta}_{(n)}} e^{-\frac{k}{2}\left(\frac{\hat{\beta}_n-\beta}{\beta}\right)^2} \left[\frac{\sin\left(\frac{\beta-\hat{\beta}_i}{2h}\right)}{\left(\frac{\beta-\hat{\beta}_i}{2h}\right)}\right]^2 d\beta}{\displaystyle\sum_{i=1}^{n} \int_{\hat{\beta}_{(1)}}^{\hat{\beta}_{(n)}} \frac{1}{\beta}\, e^{-\frac{k}{2}\left(\frac{\hat{\beta}_n-\beta}{\beta}\right)^2} \left[\frac{\sin\left(\frac{\beta-\hat{\beta}_i}{2h}\right)}{\left(\frac{\beta-\hat{\beta}_i}{2h}\right)}\right]^2 d\beta} \tag{5.15}$$

where $\hat{\beta}_{(1)}$ and $\hat{\beta}_{(n)}$ are the respective minimum and maximum values of sequence (5.13).

Since the true density function of $\hat{\beta}$, when given β, is unknown for a finite sample size, its asymptotic distribution has been used in (5.15). Such substitutions were considered in Chapter IV, and results from Monte

Carlo simulation indicated that substantial improvements in squared-error precision over maximum-likelihood could be obtained.

Since $\hat{\beta}$ is not known to be a sufficient statistic for β, we cannot write $E(\beta|\underset{\sim}{x}) = E(\beta|\hat{\beta})$; thus $\tilde{\beta}_N$ involves a further degree of approximation to the Bayes estimator. In section 5.5, results from Monte Carlo simulation indicate that $\tilde{\beta}_N$ has smaller mean-squared errors than the maximum-likelihood estimator even though $\tilde{\beta}_N$ involves these approximations.

5.3 A Smooth Empirical Bayes Estimator for β Corrected for Variance

The asymptotic distribution of $\hat{\beta}$ conditional on β was shown to be

$$\text{distr.}(\hat{\beta}|\beta) = N\left(\beta, \frac{\beta^2}{ck}\right) \tag{5.16}$$

where

$$c = \frac{\pi^2}{6}\left(\psi(2) - \log \alpha\right)^2 .$$

This distribution corresponds to the distribution given by (2.21). Hence from the calculations in section 2.5, the mean of the marginal distribution of $\hat{\beta}$ is equal to the mean of the prior distribution. Its variance, however, overestimates the prior variance by the amount

$$\frac{1}{kc}\left[\mathrm{Var}(\beta) + E^2(\beta)\right] \ .$$

We can now form the transformation

$$\beta^* = a_1\left(\hat{\beta} - E(\hat{\beta})\right) + E(\beta) \tag{5.17}$$

defined by (2.25). This transformation provides us with a new sequence of values

$$\beta_1^*, \beta_2^*, \cdots, \beta_n^* \ . \tag{5.18}$$

The marginal distribution of this sequence has a mean and a variance equivalent to those of the prior distribution. When the mean and variance of the prior distribution are known, the constant a_1 in (5.17) is given by

$$a_1 = \left(\frac{kc\ \mathrm{Var}(\beta)}{(1+kc)\ \mathrm{Var}(\beta) + E^2(\beta)}\right)^{1/2} \ . \tag{5.19}$$

Otherwise

$$a_1 = \left(\frac{1}{1+kc}\left(kc - \frac{\bar{\beta}_n^2}{s_n^2}\right)\right)^{1/2} \tag{5.20}$$

where

$$\overline{\beta}_n = \sum_{i=1}^{n} \frac{\hat{\beta}_i}{n} \tag{5.21}$$

and

$$s_n^2 = \sum_{i=1}^{n} \frac{(\hat{\beta}_i - \overline{\beta}_n)^2}{n-1} \tag{5.22}$$

Any density estimator $p_n(\beta^*)$ based on sequence (5.18) has the property that

$$\underset{\substack{n\to\infty \\ k\to\infty}}{\text{p lim}}\; p_n(\beta^*) = g(\beta) \quad .$$

Therefore when considered as a function of β, $p_n(\beta^*)$ can be used to approximate the prior density function, and a smooth estimator for β_n becomes

$$\tilde{\beta}_{N,V} = \frac{\displaystyle\sum_{i=1}^{n} \int_{\beta^*_{(1)}}^{\beta^*_{(n)}} e^{-\frac{k}{2}\left(\frac{\hat{\beta}_n - \beta}{\beta}\right)^2} \left[\frac{\sin\left(\frac{\beta-\beta_i^*}{2h}\right)}{\left(\frac{\beta-\beta_i^*}{2h}\right)}\right]^2 d\beta}{\displaystyle\sum_{i=1}^{n} \int_{\beta^*_{(1)}}^{\beta^*_{(n)}} \frac{1}{\beta}\, e^{-\frac{k}{2}\left(\frac{\hat{\beta}_n - \beta}{\beta}\right)^2} \left[\frac{\sin\left(\frac{\beta-\beta_i^*}{2h}\right)}{\left(\frac{\beta-\beta_i^*}{2h}\right)}\right]^2 d\beta} \tag{5.23}$$

where $\beta^*_{(1)}$ and $\beta^*_{(n)}$ are the respective minimum and maximum values of sequence (5.18).

5.4 Iteration of the Smooth Estimators

Consider a sequence of smooth estimates from (5.15)

$$\tilde{\beta}_{N,1},\ \tilde{\beta}_{N,2},\ \cdots,\ \tilde{\beta}_{N,n} \tag{5.24}$$

obtained in each of n previous experiments. Based on this sequence a more "precise" estimate of the realization β_n can often be determined. This estimate is obtained by replacing the prior density in the Bayes estimator by a density approximation constructed with sequence (5.24). In essence this represents an iteration of the smooth estimator $\tilde{\beta}_N$, since the kernel of integration remains unchanged and is denoted by $\tilde{\beta}'_N$. A similar iteration of the estimator $\tilde{\beta}_{N,V}$ can be obtained and will be denoted by $\tilde{\beta}'_{N,V}$. Integration in each estimator is performed over the range of values obtained from the first iteration. For example, the region of integration for $\tilde{\beta}'_N$ is from $\min(\tilde{\beta}_{N,i})$ to $\max(\tilde{\beta}_{N,j})$ where $i,j = 1,2,\cdots,n$ and $i \neq j$. When $i = 1$, we define $\tilde{\beta}_{N,1} = \hat{\beta}_1$.

5.5 Monte Carlo Simulation

The continuously smooth empirical Bayes estimators given in the previous sections were investigated for small samples by Monte Carlo simulation and compared to the maximum-likelihood estimator. In each experiment a value of β was generated from a prior distribution belonging to the Pearson family of distributions. This parameter was then used to obtain a random sample of k observations from (5.1). The maximum-likelihood and continuously smooth empirical Bayes estimators were then calculated. Five hundred repetitions of twenty experiments were made for each prior distribution, and the ratio

$$R = \frac{\text{empirical Bayes mean-squared error}}{\text{maximum-likelihood mean-squared error}}$$

was formed. As in section 3.5, it was found that the ratio R was significantly influenced by the prior distribution only through the value

$$Z = \frac{\mathrm{Var}^*(\hat{\beta}\,|\,\beta)}{\mathrm{Var}(\beta)}$$

where $*$ indicates that $E(\beta)$, the prior mean of β, has been substituted for β. In particular for a random sample of k observations from (5.1), the value of Z becomes

$$Z = \frac{E^2(\beta)\left[\frac{\pi^2}{6} + (\psi(2) - \log \alpha)^2\right]}{k\ \mathrm{Var}(\beta)} \qquad (5.25)$$

where $\psi(2) = .4227843351$ is the value of the digamma function $\psi(x)$ evaluated at $x = 2$. Since the only factors affecting the ratio R , apart from the number of experiences, are contained in (5.25), this quantity can be conveniently used to summarize and index a given situation.

As in the preceding chapters, we again wish to illustrate the robustness of the smooth estimators to the form of the prior distribution. Therefore values of β_i $(i = 1,2,\cdots,20)$ were generated from various Pearson distributions while holding the value of Z fixed at 1.5. The results are shown in Figures 26-29. In each figure the coefficients of skewness (S) and kurtosis (K) were varied to give different forms to the prior distribution. In Figure 26 the prior distribution is bell-shaped (skewed); in Figure 27, L-shaped; in Figure 28, J-shaped; and in Figure 29, U-shaped. The solid line in each figure represents the ratio $R_{N,V}$ calculated with the smooth estimator $\tilde{\beta}_{N,V}$, and the broken line represents the ratio R_N calculated with $\tilde{\beta}_N$. This representation will be used in the remaining figures of this chapter. Again we notice that the smooth estimators are quite insensitive to the form of the prior distribution as summarized by S and K . Therefore values of S and K will not be given for the remaining figures in this chapter. The values of the prior distribution are designated as follows:

$$E(\beta) = \text{prior variance of } \beta ,$$

$$V(\beta) = \text{prior mean of } \beta .$$

To determine the effect of the sample size k on the ratios $R_{N,V}$ and R_N, several runs were made with k ranging from 5 to 20. Results from these investigations revealed that the quantities in Z can vary in a manner not affecting the value of Z without having a significant effect on the ratios $R_{N,V}$ and R_N. Thus the value of Z, not the individual quantities in Z, determines the values of $R_{N,V}$ and R_N. Figures 30, 31, and 32 illustrate this point. By comparing these figures, it can be seen that, although the parameters k, α, $E(\beta)$, and $V(\beta)$ vary in each figure with the value of Z remaining 0.8, the ratios $R_{N,V}$ and R_N remain relatively unchanged. Since the ratio R remains relatively unchanged for equivalent values of Z, the individual quantities in Z will not be given for the remaining figures in this chapter.

Values of the ratios R_N and $R_{N,V}$ are plotted in Figures 33 and 34 respectively. These ratios are plotted for different values of Z ranging from 0.5 to 5.0. As in the previous chapters, we notice that as Z increases, the ratios R_N and $R_{N,V}$ decrease. In particular for a given value of Z, the values of $R_{N,V}$ are smaller than those of R_N, demonstrating the superiority of the smooth estimator $\tilde{\beta}_{N,V}$ over the smooth estimator $\tilde{\beta}_N$. As in the preceding chapters, the increase in mean-squared precision by $\tilde{\beta}_{N,V}$ is attributed to the prior density approximation. This approximation is based on a sequence of values whose marginal distribution has its mean and variance approximately equivalent to those of the prior distribution.

Figure 35 represents a typical result obtained when $\tilde{\beta}_{N,V}$ is iterated as described in section 5.4. The ratio $R'_{N,V}$ formed with the iterated smooth estimator $\tilde{\beta}'_{N,V}$ is represented by the dotted line. As in the previous chapters, we notice that the squared-error improvement achieved by using $\tilde{\beta}_{N,V}$ is slightly decreased by a second iteration. This decrease in precision is expected. The approximation used to represent the prior density in $\tilde{\beta}_{N,V}$ is based on a sequence of values whose marginal distribution has its mean and variance approximately equivalent to those of the prior distribution. The approximation used in β_N, however, is not known to have this property.

Results obtained from a second iteration of $\tilde{\beta}_N$ are presented in Figure 36 for a value of Z identical to that used in Figure 35. The ratio R'_N formed with the smooth estimator $\tilde{\beta}'_N$ is represented by the dotted line. We note that although iteration of $\tilde{\beta}_N$ does increase the mean-squared precision, this increase is not as significant as that obtained when $\tilde{\beta}_{N,V}$ is used. It is conjectured that the overestimation of the prior variance by the marginal distribution of the maximum-likelihood estimator has a far more significant effect on the prior density approximation than does the marginal variance of $\tilde{\beta}_N$. In general, iteration of the smooth estimators is discouraged. While a second iteration of $\tilde{\beta}_N$ does decrease the squared error, the decrease is not as significant as that obtained using $\tilde{\beta}_{N,V}$. A second iteration of $\tilde{\beta}_{N,V}$ usually increases the squared error.

In all cases considered in the Monte Carlo study, the estimator $\tilde{\beta}_{N,V}$ provided consistent mean-squared improvement over the maximum-likelihood estimator for two or more experiences. It was also observed to be most efficient of the smooth estimators. Since this improvement was uniform over a wide variety of Z values, we are confident that $\tilde{\beta}_{N,V}$ can be used in any situation. In case, however, some idea of the prior mean and variance is known, Figure 37 can be used to obtain some indication on the amount of improvement of $\tilde{\beta}_{N,V}$ over the maximum-likelihood estimators. In this figure, the ratio $R_{N,V}$ is plotted as a function of Z for a given number of experiences.

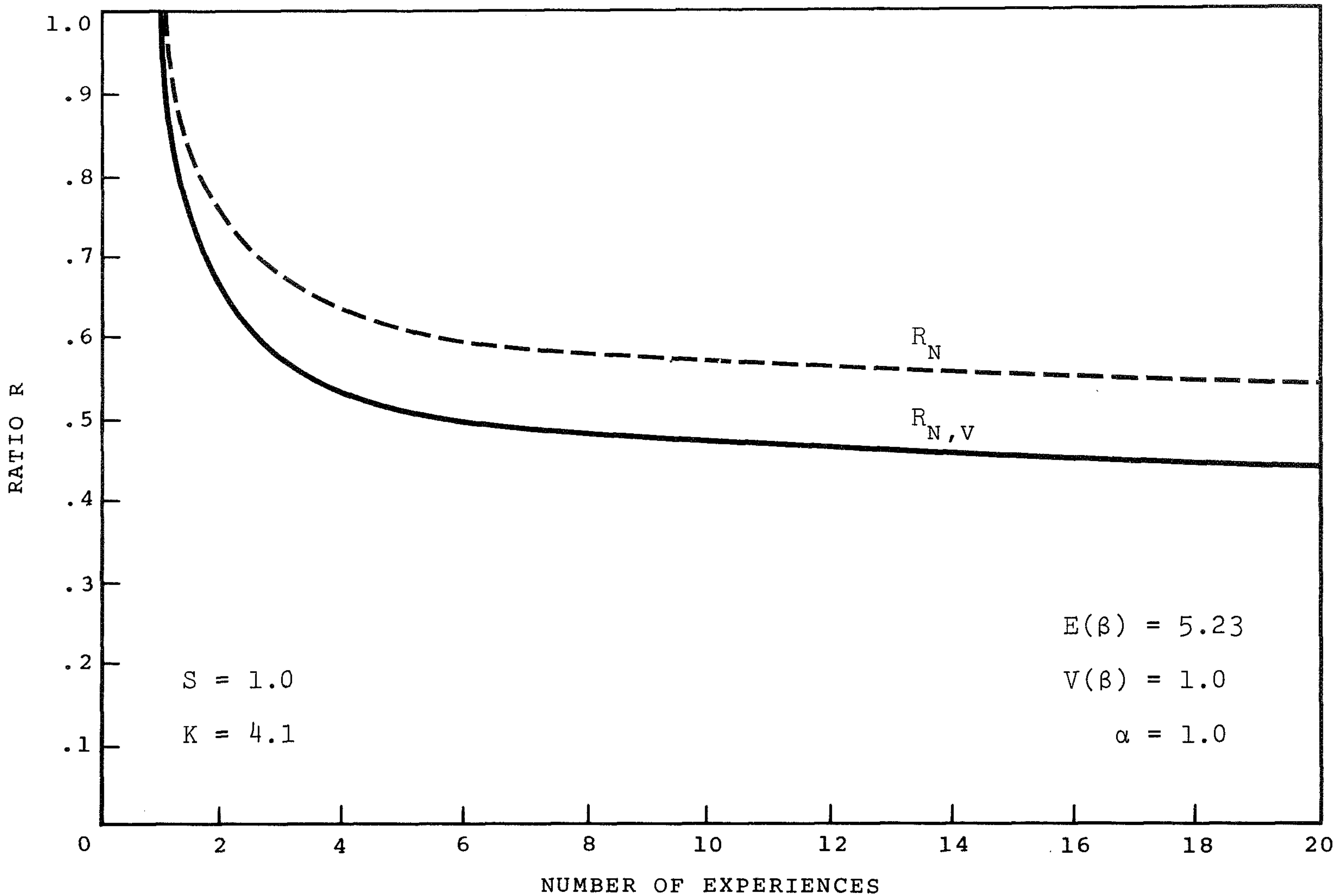

Figure 26. – R_N and $R_{N,V}$ for $Z = 1.5$; bell-shaped (skewed) prior.

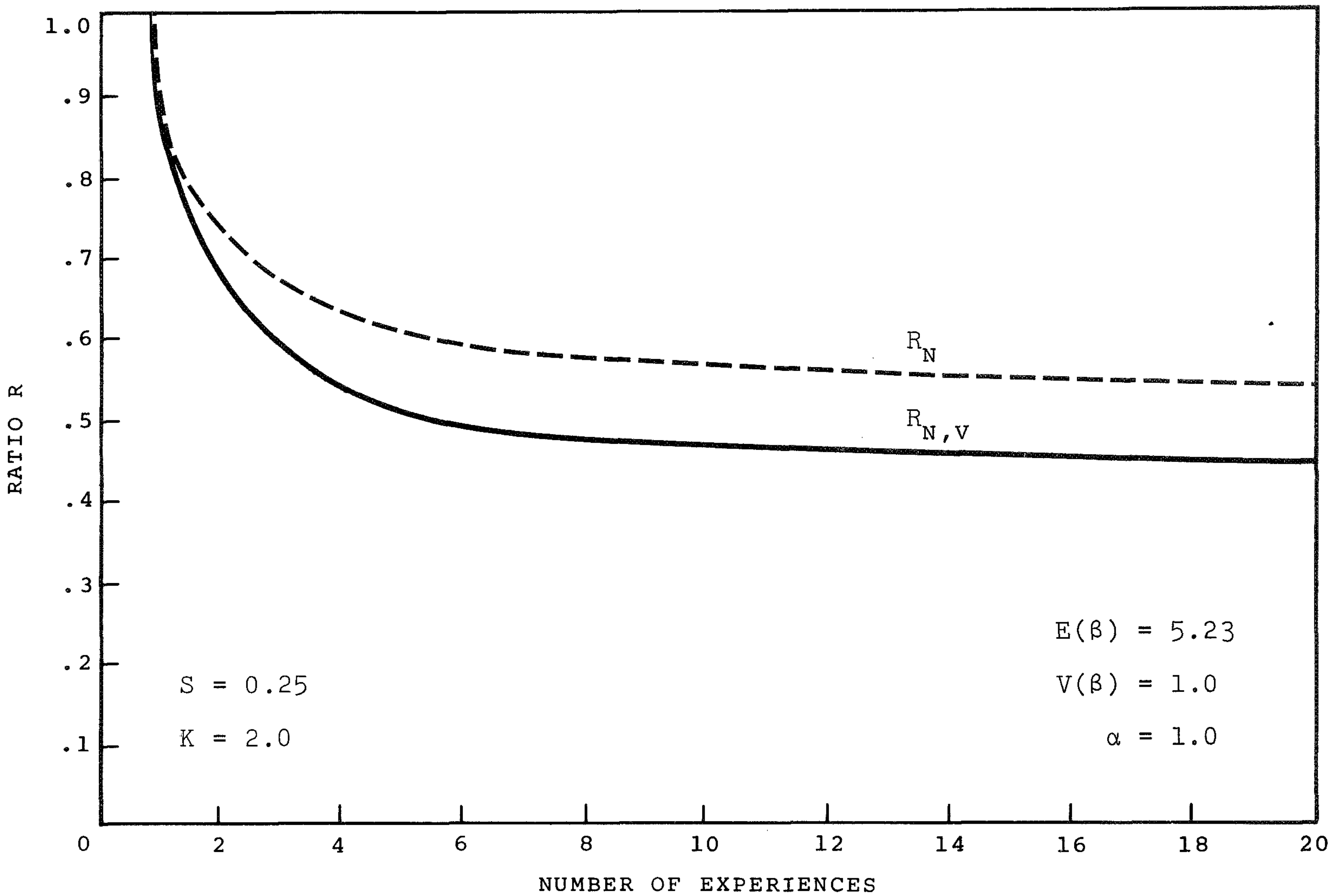

Figure 27. – R_N and $R_{N,V}$ for $Z = 1.5$; L-shaped prior.

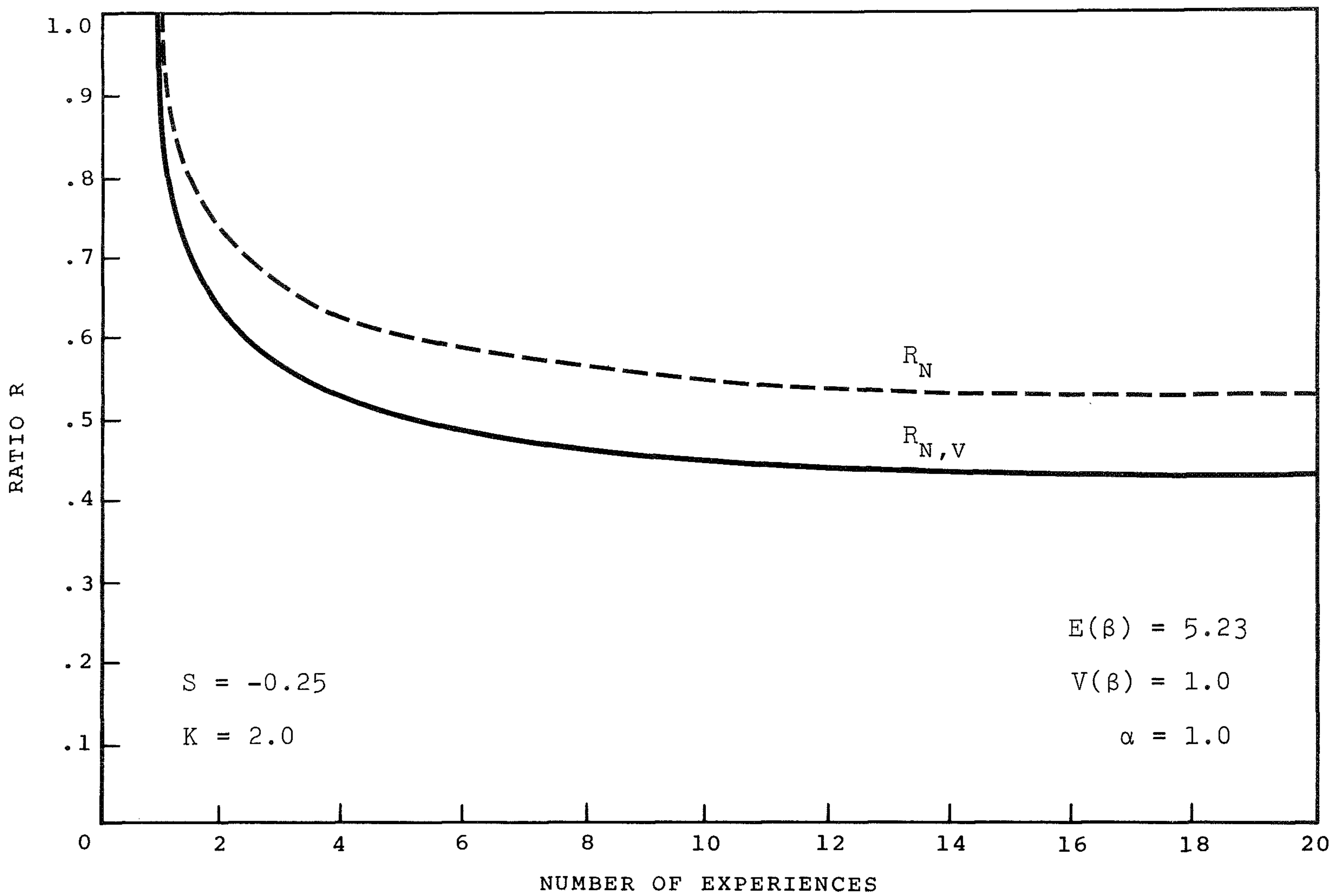

Figure 28. – R_N and $R_{N,V}$ for $Z = 1.5$; J-shaped prior.

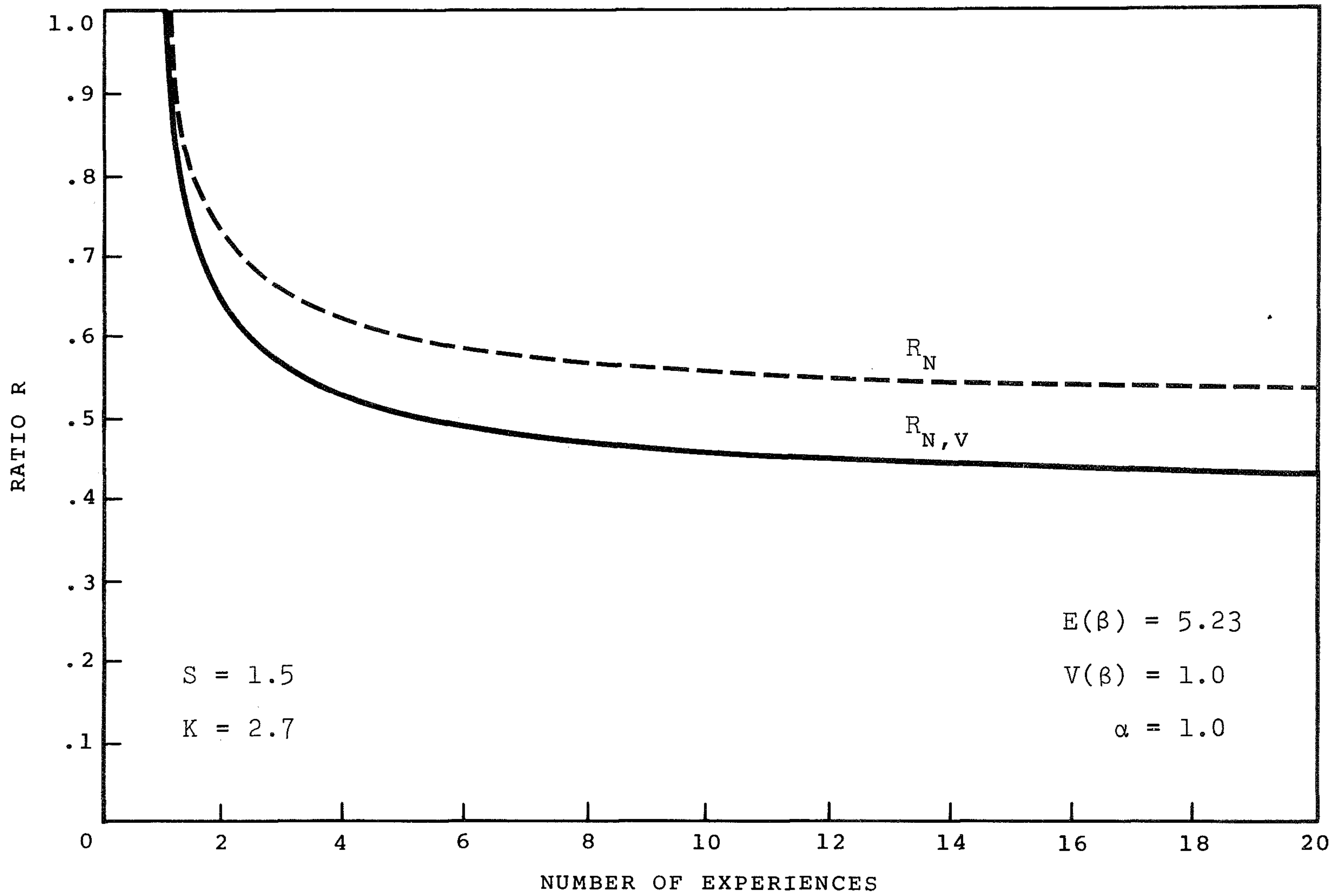

Figure 29. — R_N and $R_{N,V}$ for $Z = 1.5$; U-shaped prior.

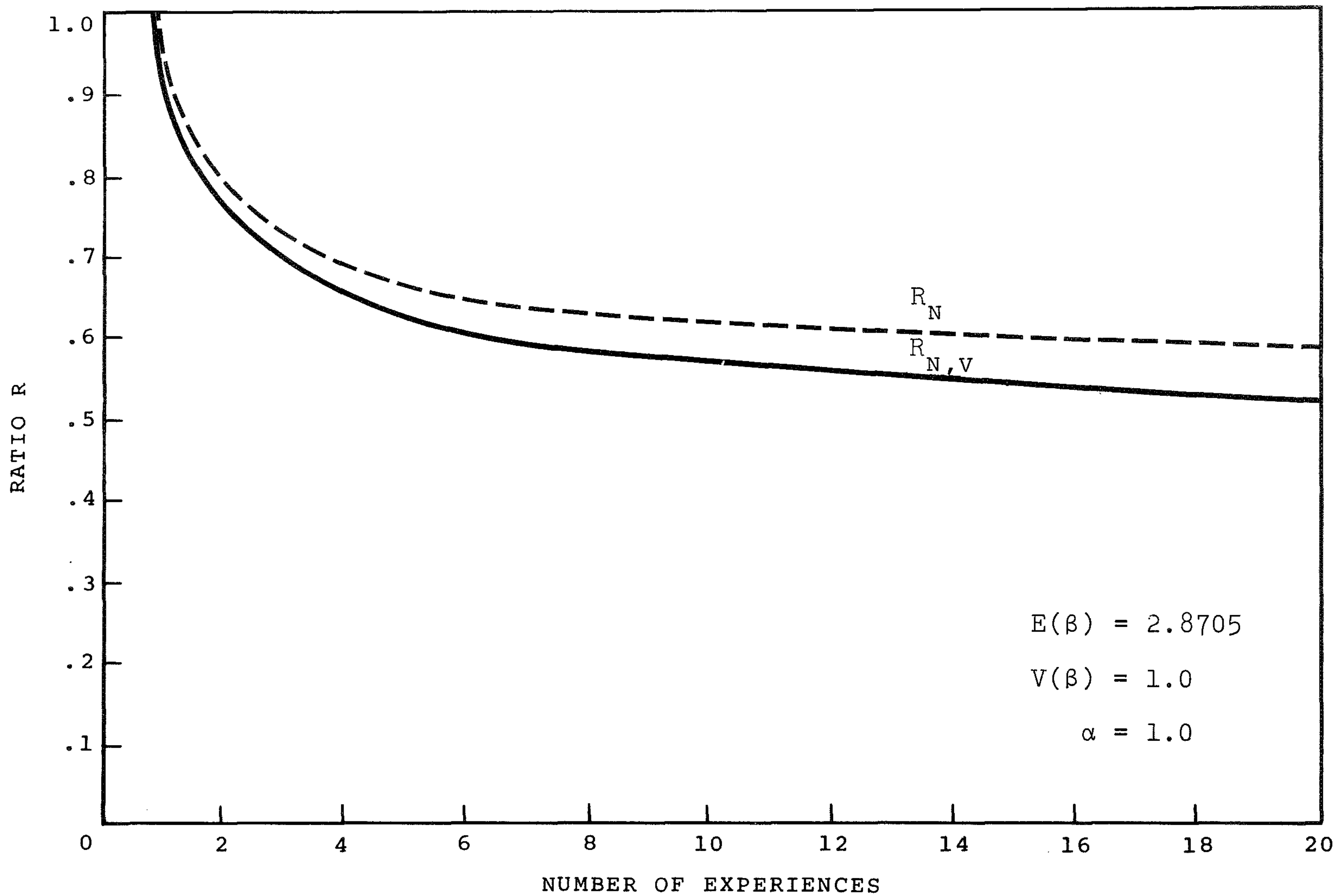

Figure 30. – R_N and $R_{N,V}$ for $Z = 0.8$ with 5 observations.

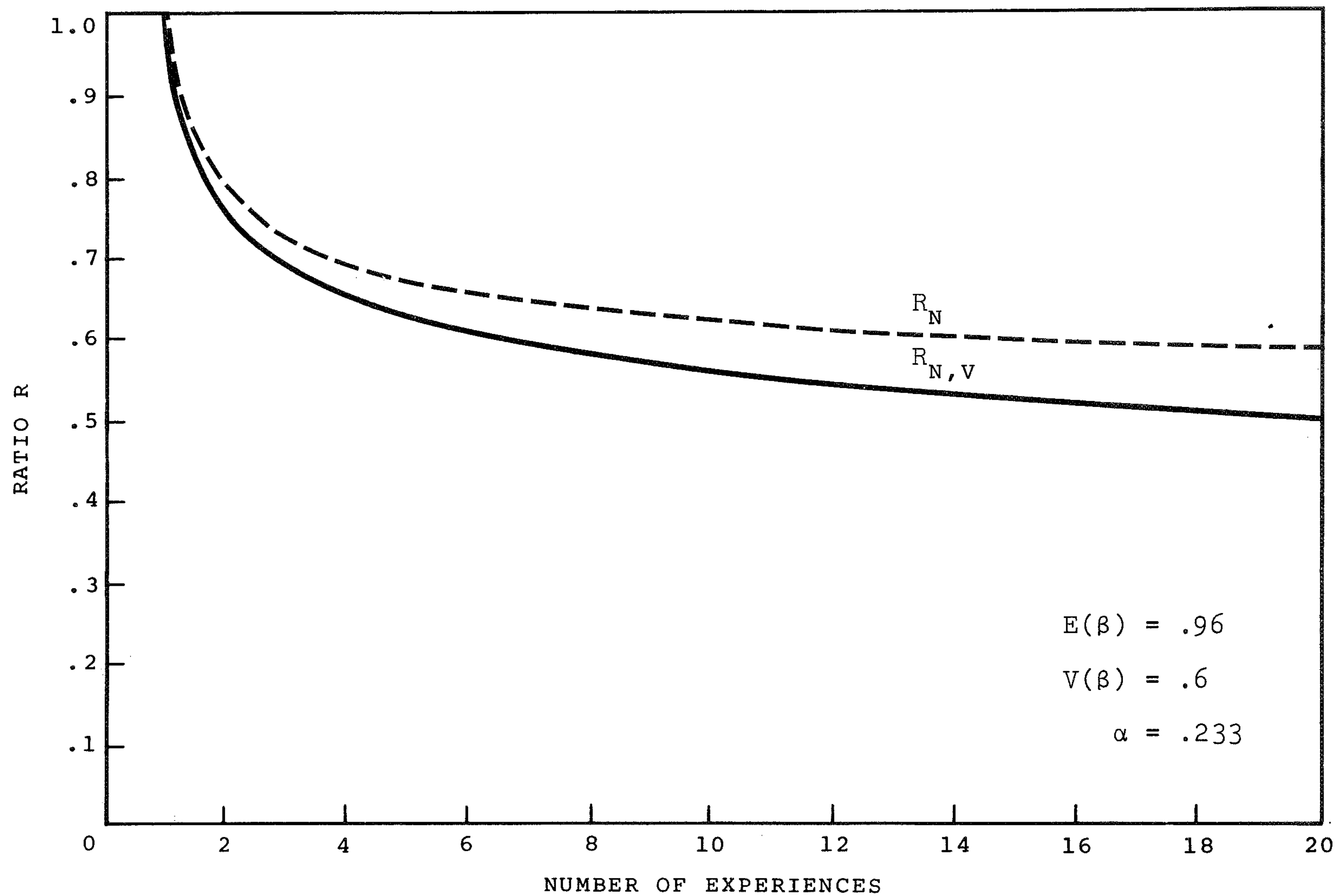

Figure 31. — R_N and $R_{N,V}$ for $Z = 0.8$ with 10 observations.

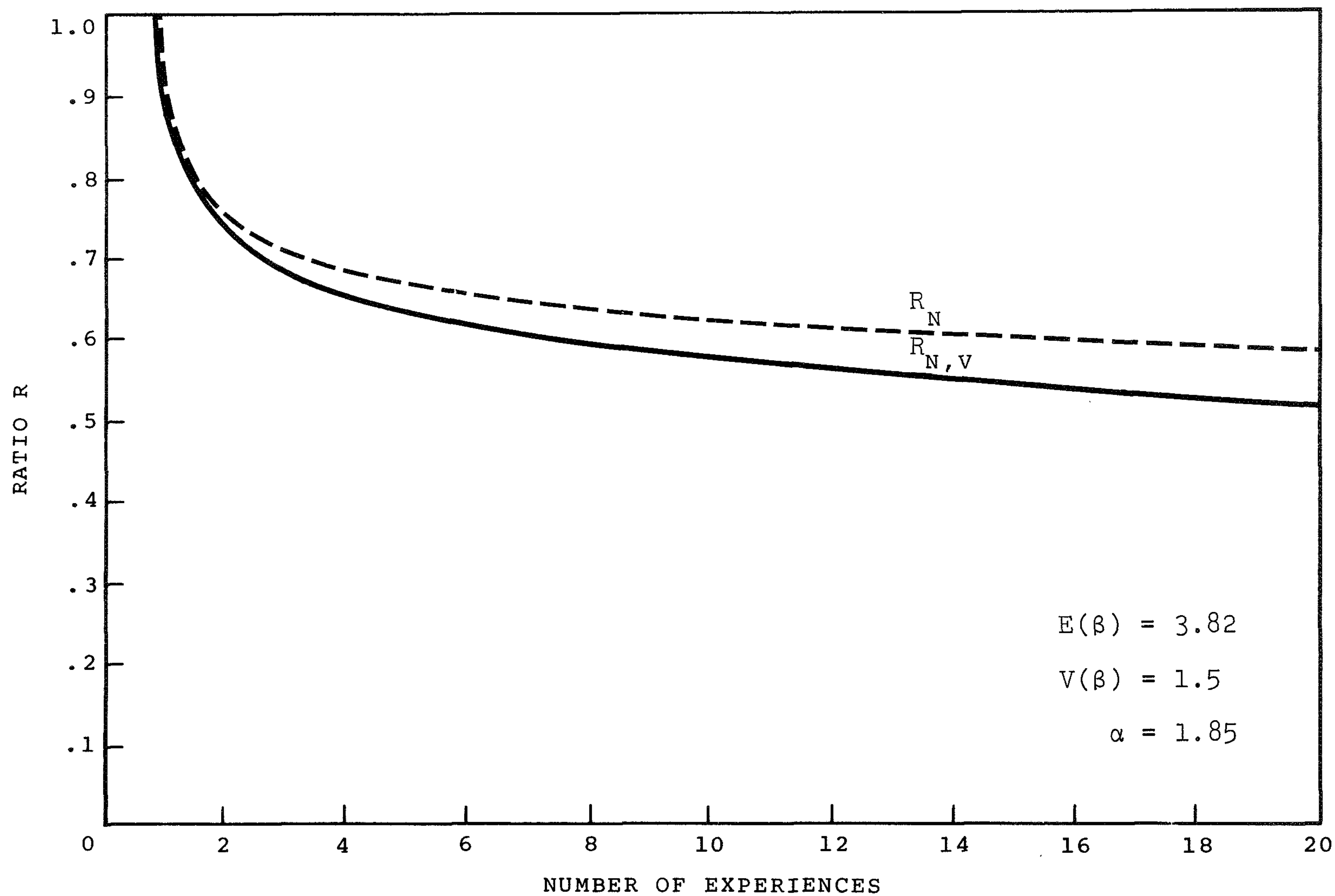

Figure 32. — R_N and $R_{N,V}$ for $Z = 0.8$ with 20 observations.

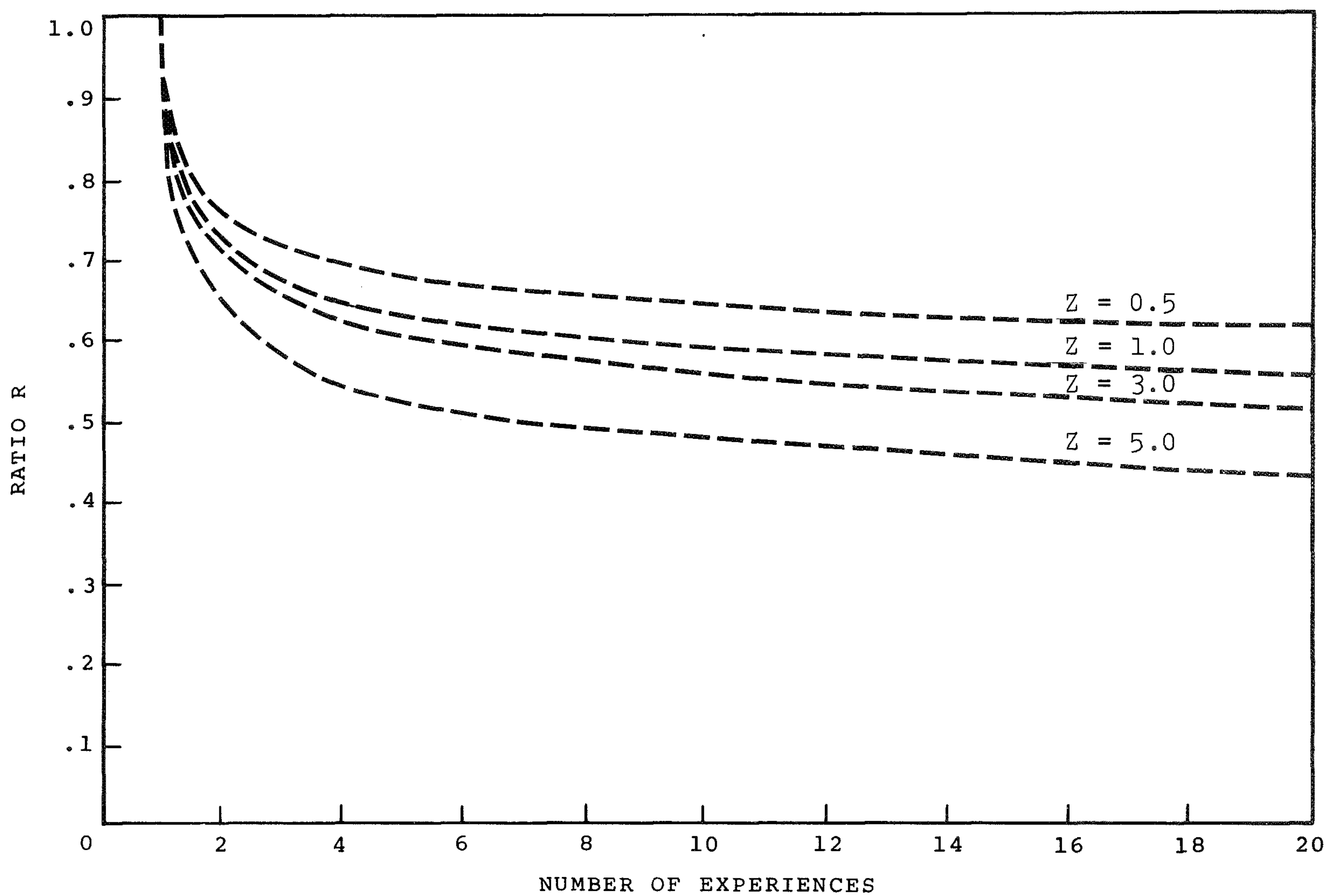

Figure 33. — R_N for several values of Z .

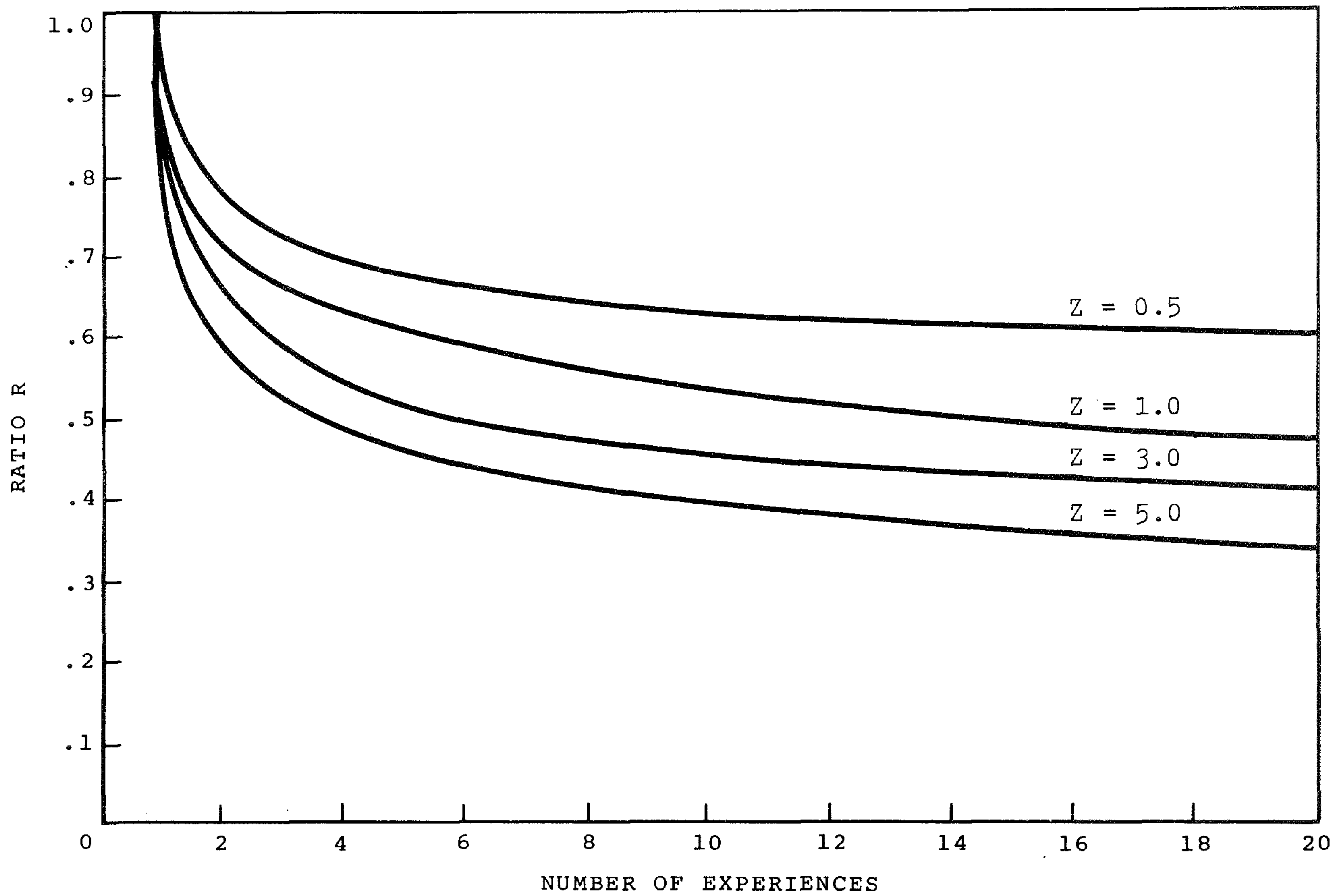

Figure 34. — $R_{N,V}$ for several values of Z .

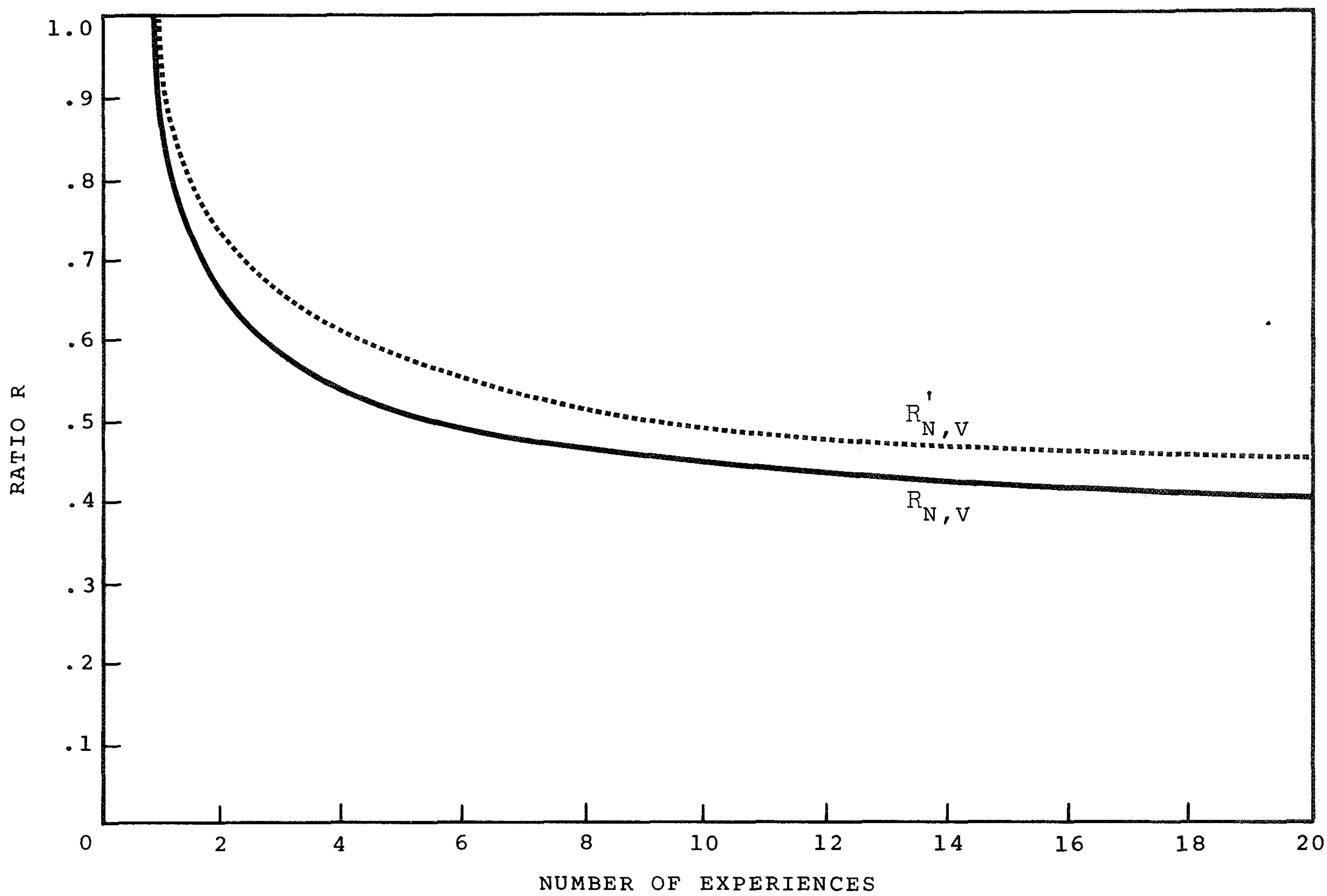

Figure 35. – $R_{N,V}$ vs $R'_{N,V}$ for $Z = 3.0$.

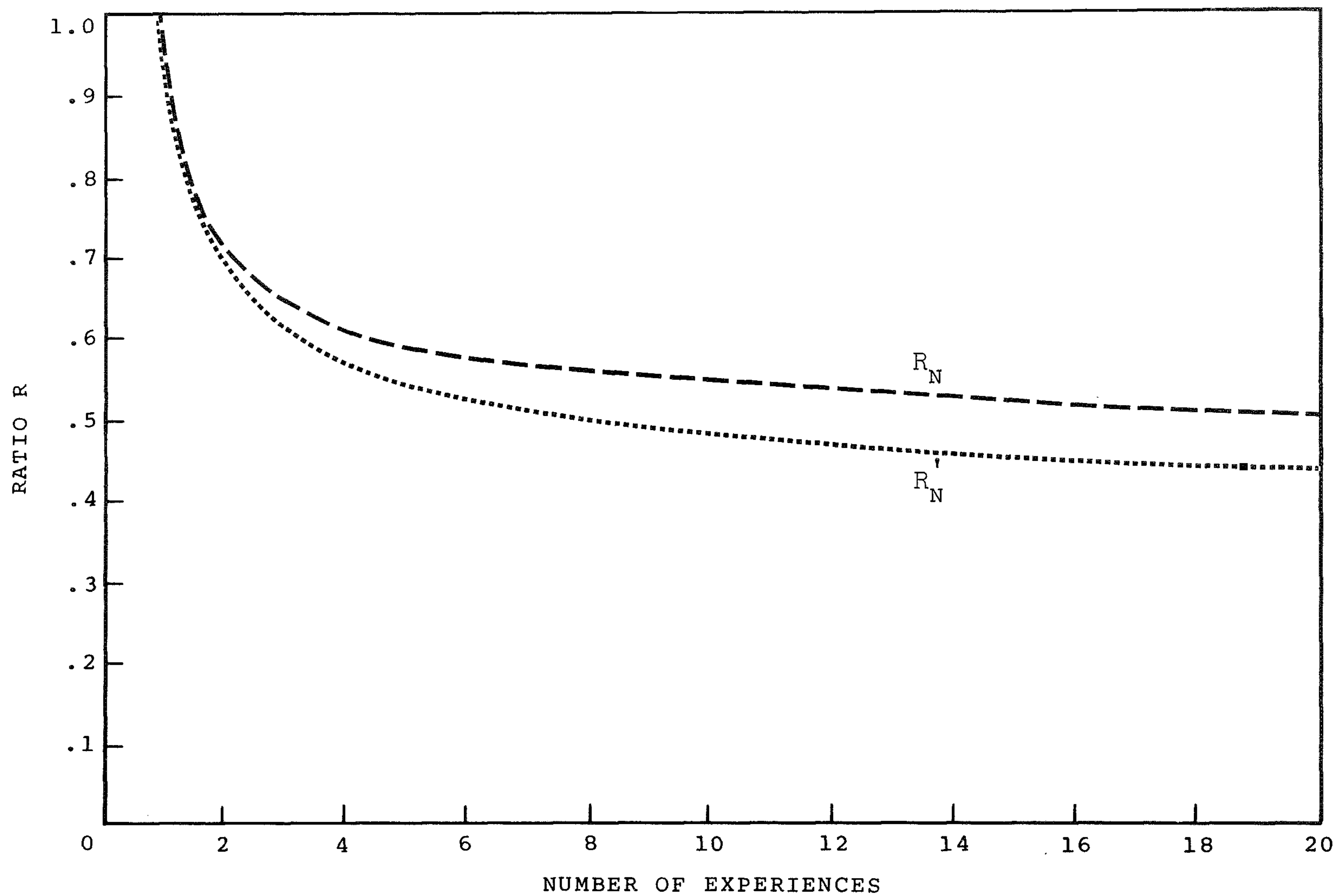

Figure 36. — R_N vs R_N' for $Z = 3.0$.

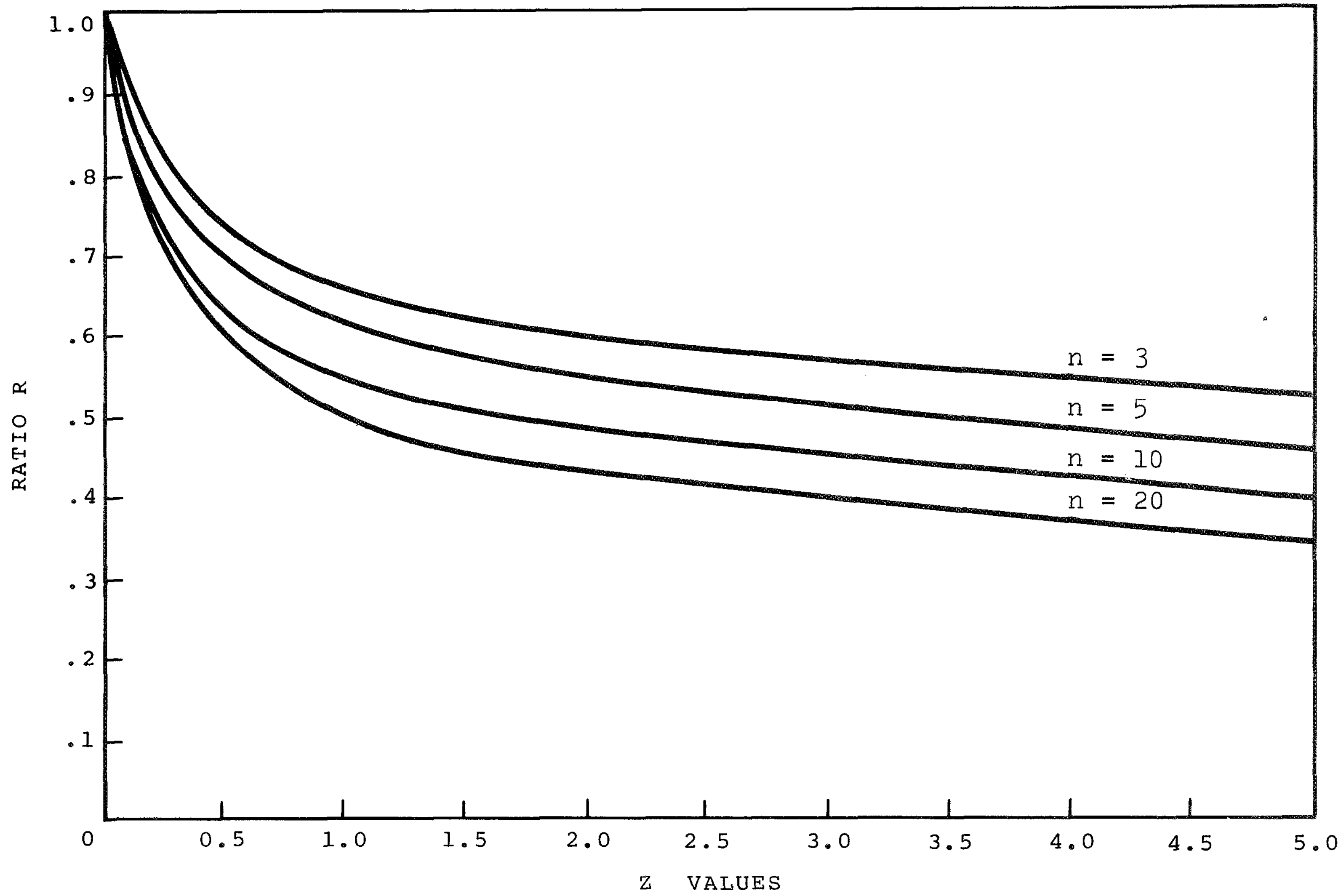

Figure 37. — $R_{N,V}$ as a function of Z for fixed number of experiences.

CHAPTER VI

ESTIMATION IN THE WEIBULL DISTRIBUTION

In this chapter, smooth empirical Bayes estimators are given for the scale parameter α and the shape parameter β in the two-parameter Weibull distribution. These estimators are proposed on the assumption that they are subject to random variation. Results from Monte Carlo simulations are reported which show that the smooth estimators have smaller squared errors than the maximum-likelihood estimators.

6.1 Maximum-Likelihood Estimation

Let X be a random variable having a Weibull distribution. If α and β , the scale and the shape parameters respectively, are random variables, then the conditional density function of X is given by

$$f(x|\alpha,\beta) = \alpha\beta x^{\beta-1}e^{-\alpha x^{\beta}} , \quad (x \geq 0;\ \alpha,\ \beta > 0) \ . \tag{6.1}$$

Consider a random sample consisting of k observations from (6.1). The likelihood function of this sample is

$$L(\underset{\sim}{x}|\beta) = (\alpha\beta)^k \prod_{i=1}^{k} x_i^{\beta-1} e^{-\alpha x_i^{\beta}} \ . \tag{6.2}$$

Taking the logarithm of (6.2), differentiating with respect to α and β in turn and equating to zero, we obtain the equations

$$\frac{\partial \log L}{\partial \alpha} = \frac{k}{\alpha} - \sum_{i=1}^{k} x_i^{\beta} = 0 \tag{6.3}$$

and

$$\frac{\partial \log L}{\partial \beta} = \frac{k}{\beta} + \sum_{i=1}^{k} \log x_i - \alpha \sum_{i=1}^{k} x_i^{\beta} \log x_i = 0 \; . \tag{6.4}$$

Eliminating α between these two equations and simplifying, we have

$$\left[\frac{\sum_{i=1}^{k} x_i^{\beta} \log x_i}{\sum_{i=1}^{k} x_i^{\beta}} - \frac{1}{\beta} \right] = \frac{1}{k} \sum_{i=1}^{k} \log x_i \tag{6.5}$$

which may now be solved to obtain the maximum-likelihood estimator $\hat{\beta}$. This can be accomplished with the aid of standard iterative procedures. In Appendix B such a procedure is described.

With $\hat{\beta}$ thus determined, α is estimated from (6.3) as

$$\hat{\alpha} = \frac{k}{\sum_{i=1}^{k} x_i^{\beta}} \tag{6.6}$$

The maximum-likelihood estimators $\hat{\alpha}$ and $\hat{\beta}$ are not known to be jointly sufficient, and their exact distributional form is unknown for small samples. The estimators are consistent and by virtue of (1.29) are distributed bivariate normal with mean vector $\underset{\sim}{\mu} = (\hat{\alpha}, \hat{\beta})$ and covariance matrix V given by

$$V = \begin{pmatrix} -E\left(\frac{\partial^2 \log L}{\partial \alpha^2}\right) & -E\left(\frac{\partial^2 \log L}{\partial \alpha \partial \beta}\right) \\ -E\left(\frac{\partial^2 \log L}{\partial \alpha \partial \beta}\right) & -E\left(\frac{\partial^2 \log L}{\partial \beta^2}\right) \end{pmatrix}^{-1} \tag{6.7}$$

Thus

$$\text{distr.}(\hat{\alpha}, \hat{\beta} | \alpha, \beta) = N_2(\underset{\sim}{\mu}, V) \quad . \tag{6.8}$$

The expected value of the mixed partial derivatives in (6.7) can be represented as

$$E\left(\frac{\partial^2 \log L}{\partial \alpha \partial \beta}\right) = - \sum_{i=1}^{k} E(x_i^{\beta} \log x_i)$$

$$= - k\, E(x^{\beta} \log x) \tag{6.9}$$

since each x_i is independently and identically distributed according to (6.1). Now

$$E(x^\beta \log x) = \int_0^\infty (x^\beta \log x)\alpha\beta x^{\beta-1} e^{-\alpha x^\beta}\, dx \qquad (6.10)$$

which becomes

$$E(x^\beta \log x) = \frac{-1}{\alpha\beta}\left(\psi(2) - \log \alpha\right) \qquad (6.11)$$

after integrating (6.10) by parts. The function $\psi(2)$ is the digamma function $\psi(x)$ evaluated at $x = 2$. Thus

$$-E\left(\frac{\partial^2 \log L}{\partial\alpha\partial\beta}\right) = \frac{k}{\alpha\beta}\left(\psi(2) - \log \alpha\right) , \qquad (6.12)$$

and the inverse of the covariance matrix can be written

$$V^{-1} = \begin{pmatrix} \frac{k}{\beta^2}\left(\frac{\pi^2}{6} + \Delta^2\right) & \frac{k\Delta}{\alpha\beta} \\ \frac{k\Delta}{\alpha\beta} & \frac{k}{\alpha^2} \end{pmatrix} \qquad (6.13)$$

where $\Delta = \psi(2) - \log \alpha$ and the main diagonal elements are given by the reciprocal of (4.9) and (5.12) respectively. The covariance matrix V can now be given as

$$V = \begin{pmatrix} \frac{\beta^2}{k}\left(\frac{\pi^2}{6}\right)^{-1} & -\frac{\alpha\beta\Delta}{k}\left(\frac{\pi^2}{6}\right)^{-1} \\ -\frac{\alpha\beta\Delta}{k}\left(\frac{\pi^2}{6}\right)^{-1} & \frac{\alpha^2}{k}\left(1 + \Delta^2\left(\frac{\pi^2}{6}\right)^{-1}\right) \end{pmatrix}. \tag{6.14}$$

6.2 Smooth Empirical Bayes Estimation

Assume that an unobservable random parameter vector $\underset{\sim}{\mu} = (\alpha, \beta)$ occurs according to the bivariate density function $g(\underset{\sim}{\mu})$. When $\underset{\sim}{\mu}$ is realized, an observable random vector $\underset{\sim}{X}$ from (6.1) occurs and a maximum-likelihood estimate of $\underset{\sim}{\mu}$ is formed. Now if this situation occurs repeatedly, then the vector-valued sequence

$$\hat{\underset{\sim}{\mu}}_1, \hat{\underset{\sim}{\mu}}_2, \cdots, \hat{\underset{\sim}{\mu}}_n \tag{6.15}$$

of maximum-likelihood estimates can be used to form a marginal density approximation $p_n(\hat{\underset{\sim}{\mu}})$. Convergence in probability of $p_n(\hat{\underset{\sim}{\mu}})$ to the prior density $g(\underset{\sim}{\mu})$ is assured by Theorem 2.2. Hence, $p_n(\hat{\underset{\sim}{\mu}})$ can be considered as a function of $\underset{\sim}{\mu}$, and a continuously smooth empirical Bayes estimate of $\underset{\sim}{\mu}_n$ can be given. Here as in Chapter V, the true density function of $\hat{\underset{\sim}{\mu}}$ given $\underset{\sim}{\mu}$ is unknown for small samples, and $\hat{\underset{\sim}{\mu}}$ is not known to be jointly sufficient for $\underset{\sim}{\mu}$. Thus $E(\underset{\sim}{\mu}|\underset{\sim}{x}) \neq E(\underset{\sim}{\mu}|\hat{\underset{\sim}{\mu}})$. Nevertheless, a smooth estimator for $\underset{\sim}{\mu}$ can be given, and in section 6.3 its ability to provide squared-error improvement over maximum-likelihood will be illustrated.

In the remainder of this chapter, it will be advantageous to omit vector notation, restricting our attention to the respective components of the vectors $\underset{\sim}{\mu}$ and $\hat{\underset{\sim}{\mu}}$. Sequence (6.15) will now be written as

$$(\hat{\alpha}_1, \hat{\beta}_1), (\hat{\alpha}_2, \hat{\beta}_2), \cdots, (\hat{\alpha}_n, \hat{\beta}_n) \ . \tag{6.16}$$

Using the multivariate density estimators proposed by Martz [19], we form

$$p_n(\hat{\alpha}, \hat{\beta}) = \frac{1}{4\pi^2 n h_\alpha h_\beta} \sum_{i=1}^{n} \left[\frac{\sin\left(\frac{\hat{\alpha}-\hat{\alpha}_i}{2h_\alpha}\right) \sin\left(\frac{\hat{\beta}-\hat{\beta}_i}{2h_\beta}\right)}{\left(\frac{\hat{\alpha}-\hat{\alpha}_i}{2h_\alpha}\right)\left(\frac{\hat{\beta}-\hat{\beta}_i}{2h_\beta}\right)} \right]^2 , \tag{6.17}$$

the marginal density estimator for $p(\hat{\alpha}, \hat{\beta})$. In a practical situation, the maximum-likelihood estimates will generally be unitized quantities. These units can be removed from the arguments of the sine function in (6.17) by defining

$$h_\alpha = s_\alpha n^{-1/5} \tag{6.18}$$

and

$$h_\beta = s_\beta n^{-1/5} \tag{6.19}$$

where

$$s_\alpha^2 = \sum_{i=1}^{n} (\hat{\alpha}_i - \bar{\alpha})^2/n \quad , \qquad s_\beta^2 = \sum_{i=1}^{n} (\hat{\beta}_i - \bar{\beta})^2/n \tag{6.20}$$

with

$$\bar{\alpha} = \sum_{i=1}^{n} \frac{\hat{\alpha}_i}{n} \quad , \qquad \bar{\beta} = \sum_{i=1}^{n} \frac{\hat{\beta}_i}{n} \tag{6.21}$$

Considering $p_n(\hat{\alpha}, \hat{\beta})$ as a function of α and β and using (6.8) as the kernel of integration, we obtain the smooth estimators

$$\tilde{\alpha}_N = \frac{\displaystyle\sum_{i=1}^{n} \int_{\hat{\beta}_{(1)}}^{\hat{\beta}_{(n)}} \int_{\hat{\alpha}_{(1)}}^{\hat{\alpha}_{(n)}} \frac{\alpha e^{-\frac{k}{2(1-\rho)^2}\Phi}}{\sigma_\alpha \sigma_\beta \sqrt{1-\rho}} \, p_n(\alpha, \beta)\, d\alpha\, d\beta}{\displaystyle\sum_{i=1}^{n} \int_{\hat{\beta}_{(1)}}^{\hat{\beta}_{(n)}} \int_{\hat{\alpha}_{(1)}}^{\hat{\alpha}_{(n)}} \frac{e^{-\frac{k}{2(1-\rho)^2}\Phi}}{\sigma_\alpha \sigma_\beta \sqrt{1-\rho}} \, p_n(\alpha, \beta)\, d\alpha\, d\beta} \tag{6.22}$$

and

$$\tilde{\beta}_N = \frac{\displaystyle\sum_{i=1}^{n} \int_{\hat{\beta}_{(1)}}^{\hat{\beta}_{(n)}} \int_{\hat{\alpha}_{(1)}}^{\hat{\alpha}_{(n)}} \frac{\beta e^{-\frac{k}{2(1-\rho)^2}\Phi}}{\sigma_\alpha \sigma_\beta \sqrt{1-\rho}} \, p_n(\alpha, \beta)\, d\alpha\, d\beta}{\displaystyle\sum_{i=1}^{n} \int_{\hat{\beta}_{(1)}}^{\hat{\beta}_{(n)}} \int_{\hat{\alpha}_{(1)}}^{\hat{\alpha}_{(n)}} \frac{e^{-\frac{k}{2(1-\rho)^2}\Phi}}{\sigma_\alpha \sigma_\beta \sqrt{1-\rho}} \, p_n(\alpha, \beta)\, d\alpha\, d\beta} \tag{6.23}$$

for α_n and β_n respectively. Here

$$\sigma_\alpha = \frac{\alpha}{\sqrt{k}} \left(1 + \Delta^2 \left(\frac{\pi^2}{6}\right)^{-1}\right)^{1/2}, \qquad \sigma_\beta = \frac{\beta}{\sqrt{k}} \left(\frac{\pi^2}{6}\right)^{-1/2}$$

$$\rho = \frac{-\Delta}{\left(\frac{\pi^2}{6} + \Delta^2\right)^{1/2}}, \qquad \Phi = \xi^2 - 2\rho\xi\eta + \eta^2$$

where

$$\Delta = \psi(2) - \log \alpha, \qquad \xi = \frac{\hat{\alpha}_n - \alpha}{\alpha\left(1 + \Delta^2\left(\frac{\pi^2}{6}\right)^{-1}\right)^{1/2}},$$

$$\eta = \frac{\hat{\beta}_n - \beta}{\beta\left(\frac{\pi^2}{6}\right)^{-1}}; \quad \text{and}$$

$$p_n(\alpha, \beta) = \frac{1}{2\pi n h_\alpha h_\beta} \sum_{i=1}^{n} \left[\frac{\sin\left(\frac{\alpha-\hat{\alpha}_i}{2h_\alpha}\right) \sin\left(\frac{\beta-\hat{\beta}_i}{2h_\beta}\right)}{\left(\frac{\alpha-\hat{\alpha}_i}{2h_\alpha}\right)\left(\frac{\beta-\hat{\beta}_i}{2h_\beta}\right)} \right]^2 \tag{6.24}$$

where h_α and h_β are given by (6.18) and (6.19) respectively. The limits of integration in (6.22) were obtained by ordering the components of sequence (6.16) for $\hat{\alpha}$ and $\hat{\beta}$ respectively.

The smooth estimators $\tilde{\alpha}_N$ and $\tilde{\beta}_N$ require the estimation of bivariate density functions which are difficult to estimate accurately. Also, to obtain point estimates from each of these estimators, double integration must be performed over regions which have been approximated from sample data. To avoid a substantial decrease in mean-squared precision which could be created by the compounding effect of such approximations, we propose the use of marginal empirical Bayes estimators as a possible alternative.

The smooth empirical Bayes estimator $\tilde{\alpha}_N$ has been used to represent an approximation to $E(\alpha|\hat{\alpha}, \hat{\beta})$, even though $\hat{\alpha}$ and $\hat{\beta}$ are not jointly sufficient for α and β . We extend this approximation further by constructing a continuously smooth estimator for α based on the marginal Bayes estimator $E(\alpha|\hat{\alpha})$. Similarly, a continuously smooth estimator for β will be based on the marginal Bayes estimator $E(\beta|\hat{\beta})$. By constructing such estimators, we are tacitly assuming that $\hat{\alpha}$ and $\hat{\beta}$ are independently distributed and marginally sufficient for α and β respectively, and that α and β are independently distributed. The results of such approximations are considered in section 6.3. We remark that the use of marginal empirical Bayes estimators is partly motivated by the results Clemmer and Krutchkoff [3] obtained with such approximations.

The marginal empirical Bayes estimator for α can therefore be written as

$$\tilde{\alpha}_M = \frac{\sum_{i=1}^{n} \int_{\hat{\alpha}_{(1)}}^{\hat{\alpha}_{(n)}} \frac{\alpha}{\sigma_\alpha} e^{-\frac{1}{2}\left(\frac{\hat{\alpha}_n - \alpha}{\sigma_\alpha}\right)^2} p_n(\alpha)\, d\alpha}{\sum_{i=1}^{n} \int_{\hat{\alpha}_{(1)}}^{\hat{\alpha}_{(n)}} \frac{1}{\sigma_\alpha} e^{-\frac{1}{2}\left(\frac{\hat{\alpha}_n - \alpha}{\sigma_\alpha}\right)^2} \cdot p_n(\alpha)\, d\alpha} \tag{6.25}$$

where $p_n(\alpha)$ is given by

$$p_n(\alpha) = \frac{1}{2\pi n h_\alpha} \sum_{i=1}^{n} \left[\frac{\sin\left(\frac{\alpha - \hat{\alpha}_i}{2h_\alpha}\right)}{\left(\frac{\alpha - \hat{\alpha}_i}{2h_\alpha}\right)} \right]^2 \qquad (6.26)$$

The marginal empirical Bayes estimator for β is given by

$$\tilde{\beta}_M = \frac{\displaystyle\sum_{i=1}^{n} \int_{\hat{\beta}_{(1)}}^{\hat{\beta}_{(n)}} \frac{\beta}{\sigma_\beta} e^{-\frac{1}{2}\left(\frac{\hat{\beta}_n - \beta}{\sigma_\beta}\right)^2} p_n(\beta)\, d\beta}{\displaystyle\sum_{i=1}^{n} \int_{\hat{\beta}_{(1)}}^{\hat{\beta}_{(n)}} \frac{1}{\sigma_\beta} e^{-\frac{1}{2}\left(\frac{\hat{\beta}_n - \beta}{\sigma_\beta}\right)^2} p_n(\beta)\, d\beta} \qquad (6.27)$$

where $p_n(\beta)$ is given by

$$p_n(\beta) = \frac{1}{2\pi n h_\beta} \sum_{i=1}^{n} \left[\frac{\sin\left(\frac{\beta - \hat{\beta}_i}{2h_\beta}\right)}{\left(\frac{\beta - \hat{\beta}_i}{2h_\beta}\right)} \right]^2 \qquad (6.28)$$

6.3 Monte Carlo Simulation

The continuously smooth empirical Bayes estimators $\tilde{\alpha}$ and $\tilde{\beta}$ were investigated by Monte Carlo simulation and compared to the maximum-likelihood estimators. The criterion for comparison was mean-squared error, and therefore the ratio

$$R = \frac{\text{empirical Bayes mean-squared error}}{\text{maximum-likelihood mean-squared error}}$$

was of interest. Here the notation $\tilde{\alpha}$ and $\tilde{\beta}$ is used to represent any of the smooth estimators given in the preceding section for the parameters α and β respectively.

Values of α and β were generated from chosen prior distributions. Then a random sample of k observations corresponding to the realizations α and β were generated by (6.1). The maximum-likelihood estimators $\hat{\alpha}$ and $\hat{\beta}$ were found, and their squared deviations from the values of the corresponding parameters were calculated. For the second experiment new values of α and β were generated. The procedure was repeated with $\hat{\alpha}$ and $\hat{\beta}$, and their squared deviations were calculated. For this experiment, $\tilde{\alpha}$ and $\tilde{\beta}$ and their squared deviations were also calculated. This procedure was repeated 20 times, and each time $\tilde{\alpha}$ and $\tilde{\beta}$ were calculated using the present values of $\hat{\alpha}$ and $\hat{\beta}$ as well as all previous maximum-likelihood estimates. Five hundred repetitions of this run of 20 experiments were then made, and the average of the squared deviations of $\hat{\alpha}$, $\hat{\beta}$, $\tilde{\alpha}$, and $\tilde{\beta}$ were formed as estimates of $E(\hat{\alpha} - \alpha)^2$, $E(\hat{\beta} - \beta)^2$, $E(\tilde{\alpha} - \alpha)^2$ and $E(\tilde{\beta} - \beta)^2$ respectively. Then the ratios R_α and R_β were calculated utilizing these estimated mean-squared errors.

In the Monte Carlo simulation, numerical integration was performed using the Gauss quadrature formula described in Appendix A. The numerical integrations in the smooth estimators $\tilde{\alpha}_M$ and $\tilde{\beta}_M$ were calculated to a desired degree of accuracy by means of halving the intervals of integration as discussed in Appendix A. The numerical solution of several double integrals is required to form the smooth estimators $\tilde{\alpha}_N$ and $\tilde{\beta}_N$. Therefore in the Monte Carlo simulation, checking for convergence proved to be extremely time consuming. Several runs were made in which the integrals were calculated without requiring the time-consuming accuracy check. The results from these runs were directly compared to those obtained when accuracy checking was employed. This comparison showed that the results almost always were comparable, to three significant digits, and in cases where they differed, the mean-squared errors of $\tilde{\alpha}_N$ and $\tilde{\beta}_N$, formed without consideration of integral convergence, were only slightly greater than those obtained when integral convergence was considered. Thus any bias associated with the ratios R_α or R_β formed with these mean-squared errors would be in favor of the maximum-likelihood estimators. Since this bias was slight and computer run time was significantly reduced, any gain in precision by checking for accuracy was sacrificed in favor of the reduced run time.

The parameters α and β were generated independently in the Monte Carlo studies. The theory does not require that α and β be independently distributed; however for ease in obtaining random values from the Pearson distributions this was taken to be the case. As in the previous chapters, results indicated that the ratio R depended on the distributions only as they influenced the value of the ratio of the conditional variance of the maximum-likelihood estimator to the variance of the parameter, given the corresponding parameter value. Thus, for α

$$Z_\alpha = \frac{E^2(\alpha)\left(1 + \Delta^2 \left(\frac{\pi^2}{6}\right)^{-1}\right)}{k \operatorname{Var}(\alpha)} \qquad (6.29)$$

where $\Delta = \psi(2) - \log E(\alpha)$ and $\psi(2) = .4227843351$; and, for β ,

$$Z_\beta = \frac{E^2(\beta)\left(\frac{\pi^2}{6}\right)^{-1}}{k \ \mathrm{Var}(\beta)} \quad . \tag{6.30}$$

In (6.29) and (6.30), the values of the parameters α and β have been replaced by their expected values. Apart from the number of experiences, the only factors found to affect the ratios R_α and R_β were contained in Z_α and Z_β respectively. Therefore, these quantities can be conveniently used to summarize and index a given situation. In particular, it was found that for a given value of Z , the ratio R remained relatively unchanged regardless of the correlation between the maximum-likelihood estimators.

In order to support the claim that the smooth estimators are indeed robust to the form of the prior distribution, the parameters α and β were independently generated from various Pearson distributions. For all types of Pearson prior distributions with varying coefficients of skewness and kurtosis, the ratio R_α and R_β have been observed to vary only slightly for a given number of experiences, providing the values of Z_α and Z_β remain unchanged. Illustrations of this fact are presented in Figures 38-41 for given values of $Z_\alpha = Z_\beta = 2.0$. For convenience the same coefficients of skewness (S) and kurtosis (K) were used in each situation; however the first four moments of the prior distribution were different. In Figures 38, 39, 40, and 41, the prior distribution is bell-shaped (skewed), L-shaped, J-shaped, and U-shaped respectively. The solid line represents the ratio $R_{\alpha,N}$ calculated with the smooth estimator $\tilde{\alpha}_N$ given by (6.22), and the broken line represents the ratio $R_{\beta,N}$ calculated with the smooth estimator $\tilde{\alpha}_\beta$ given by (6.23). The parameters of the prior distributions are designated as follows:

$E(\alpha)$ = prior mean of α

$V(\alpha)$ = prior variance of α

$E(\beta)$ = prior mean of β

$V(\beta)$ = prior variance of β

We remark that the ratios $R_{\alpha,M}$ and $R_{\beta,M}$ calculated with the marginal empirical Bayes estimators $\tilde{\alpha}_M$ and $\tilde{\beta}_M$ given by (6.25) and (6.27) respectively are also robust to the form of prior distribution.

In the Monte Carlo study it has been repeatedly observed that the maximum-likelihood estimates vary widely for a random sample consisting of less than 15 observations. These fluctuations caused the averaged squared errors over 500 repetitions to represent poor approximations to $E(\hat{\alpha}-\alpha)^2$ and $E(\hat{\beta}-\beta)^2$. Therefore in each case reported in this section, k was fixed at 20. For $k \geq 15$ we observed that for a given value of Z, the ratio R was unaffected by choice of sample size. Thus restricting k to 20, causes no loss in generality.

The results given above were based on the assumption of independent prior distributions for α and β. A question which naturally arises is: What effect does correlation between α and β have on the ratios R_α and R_β? To answer this question a bivariate normal prior distribution was assumed for the parameters α and β. For given values of Z_α and Z_β it was found that as the correlation ρ between α and β increased, the ratios $R_{\alpha,N}$ and $R_{\beta,N}$ decreased. Also the same degree of positive and negative correlation gave similar results. For example, correlations of $\rho = 0.5$ and $\rho = -0.5$ gave similar results for equivalent values of Z_α and Z_β.

Figures 42-47 illustrate the increase in mean-squared precision achieved by the smooth estimators $\tilde{\alpha}_N$

and $\tilde{\beta}_N$ over the maximum-likelihood estimators $\hat{\alpha}$ and $\hat{\beta}$ for various values of ρ and Z. Figures 42, 43, and 44 illustrate the improvements achieved by $\tilde{\alpha}_N$ when $\rho = 0.0$, $\rho = 0.5$, and $\rho = 0.9$ respectively for several values of Z_α. Figures 45, 46, and 47 illustrate the improvement achieved by $\tilde{\beta}_N$ when $\rho = 0.0$, $\rho = 0.5$, and $\rho = 0.9$ respectively for several values of Z_β. We note that for a given value of Z_α, as ρ increases, $R_{\alpha,N}$ decreases. Similarly for a given value of Z_β, as ρ increases, $R_{\beta,N}$ decreases.

The amount of correlation between α and β has been observed to have no effect on the ratios $R_{\alpha,M}$ and $R_{\beta,M}$ formed with the marginal estimators $\tilde{\alpha}_M$ and $\tilde{\beta}_M$ respectively. This is, of course, expected since the smooth estimators were based on the assumption of independent prior distributions. Values of the ratios $R_{\alpha,M}$ and $R_{\beta,M}$ are plotted in Figures 48 and 49 respectively for various values of Z_α and Z_β ranging from 0.5 to 5.0. It is of particular importance that for a given value of Z_α and a fixed number of experiences, the value of $R_{\alpha,M}$ is less than that of $R_{\alpha,N}$ regardless of the value of ρ. Similar results are witnessed by observing corresponding values of $R_{\beta,M}$ and $R_{\beta,N}$. Thus the marginal smooth estimators $\tilde{\alpha}_M$ and $\tilde{\beta}_M$ provide "better" results than the estimators

$\tilde{\alpha}_N$ and $\tilde{\beta}_N$ even though the former are based on the assumption of independent prior densities.

It is conjectured that the compounding effect of being unable to accurately approximate bivariate densities and having to perform double integration over regions approximated from sample data cause this phenomenon. The errors produced by such approximations in $\tilde{\alpha}_N$ and $\tilde{\beta}_N$ appear to be far more significant than the errors produced by the independent assumptions on which the marginal estimators $\tilde{\alpha}_M$ and $\tilde{\beta}_M$ are based. We are confident that $\tilde{\alpha}_M$ and $\tilde{\beta}_M$ will give "best" results in any situation since they provided uniform improvement over $\tilde{\alpha}_N$ and $\tilde{\beta}_N$ regardless of value of Z or the degree of correlation between α and β .

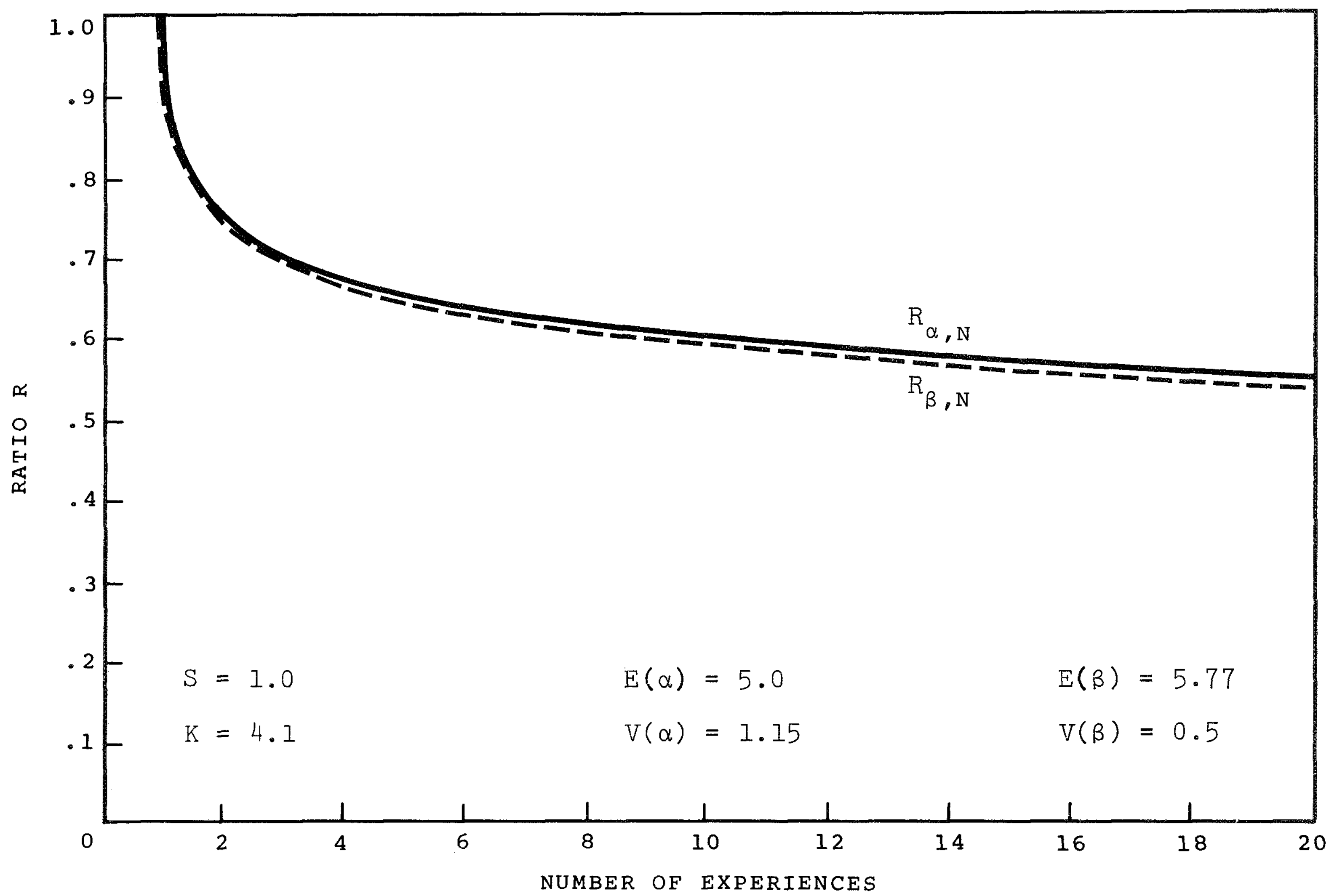

Figure 38. – $R_{\alpha,N}$ and $R_{\beta,N}$ for $Z = 2.0$; bell-shaped (skewed) prior.

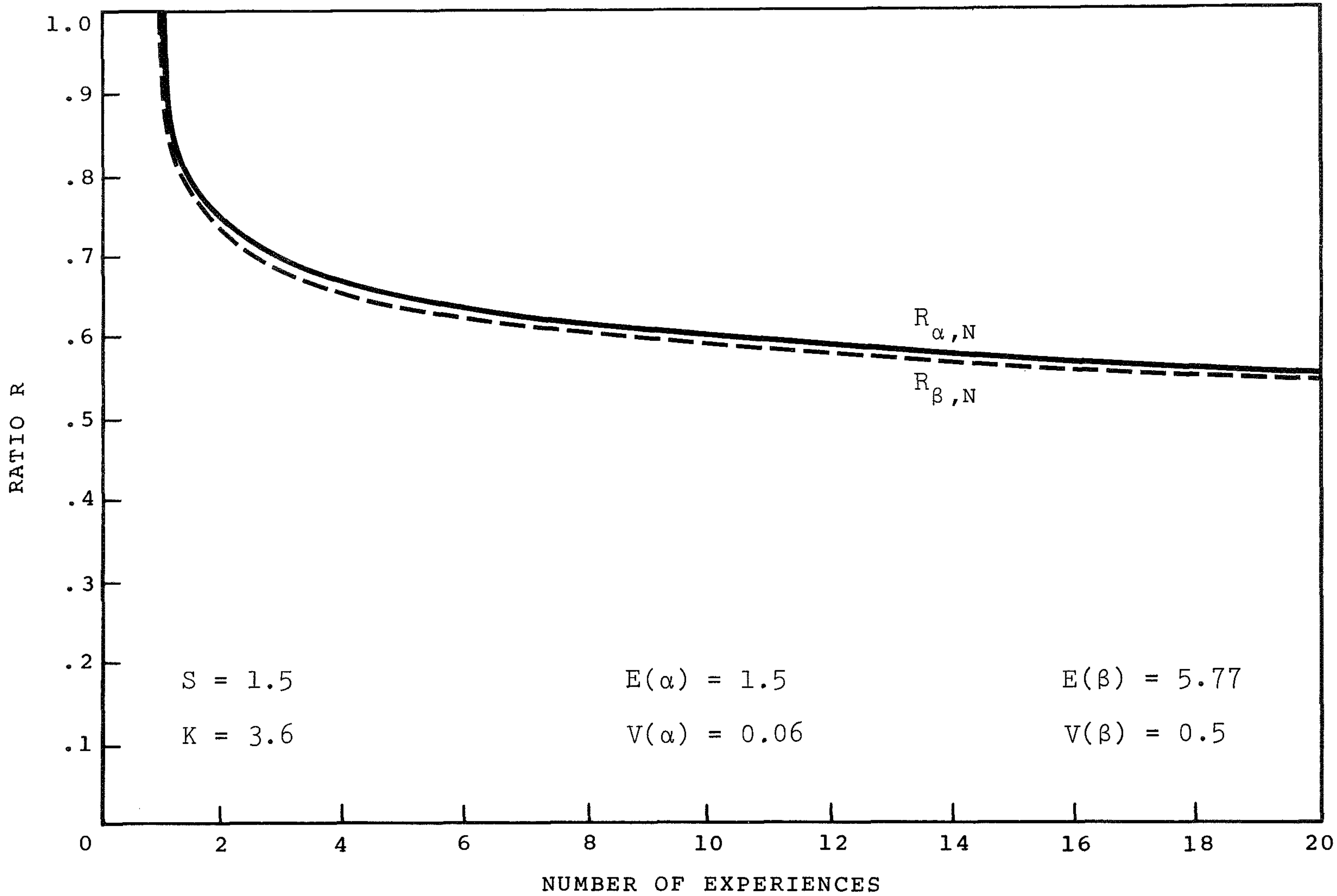

Figure 39. — $R_{\alpha,N}$ and $R_{\beta,N}$ for $Z = 2.0$; L-shaped prior.

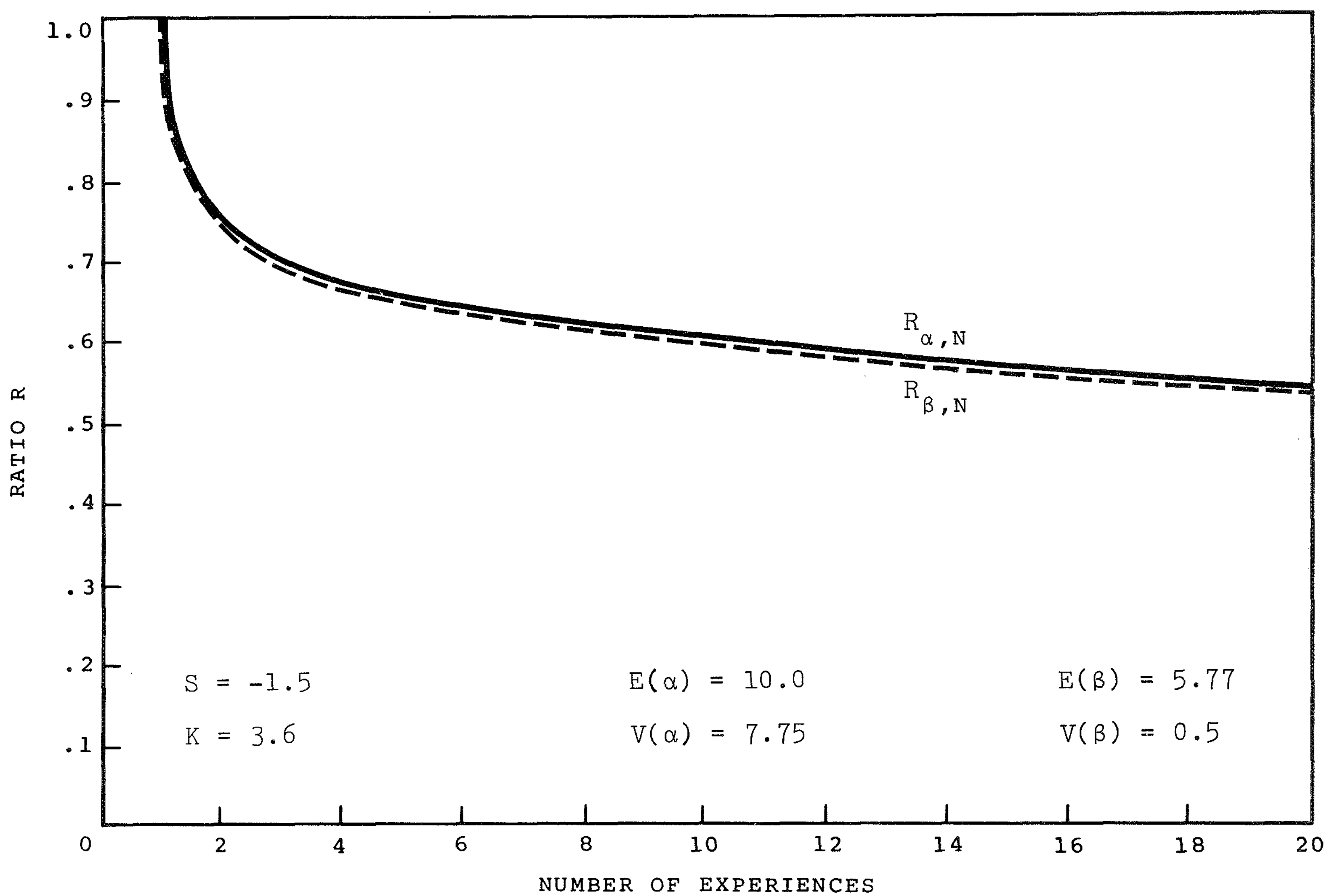

Figure 40. — $R_{\alpha,N}$ and $R_{\beta,N}$ for $Z = 2.0$; J-shaped prior.

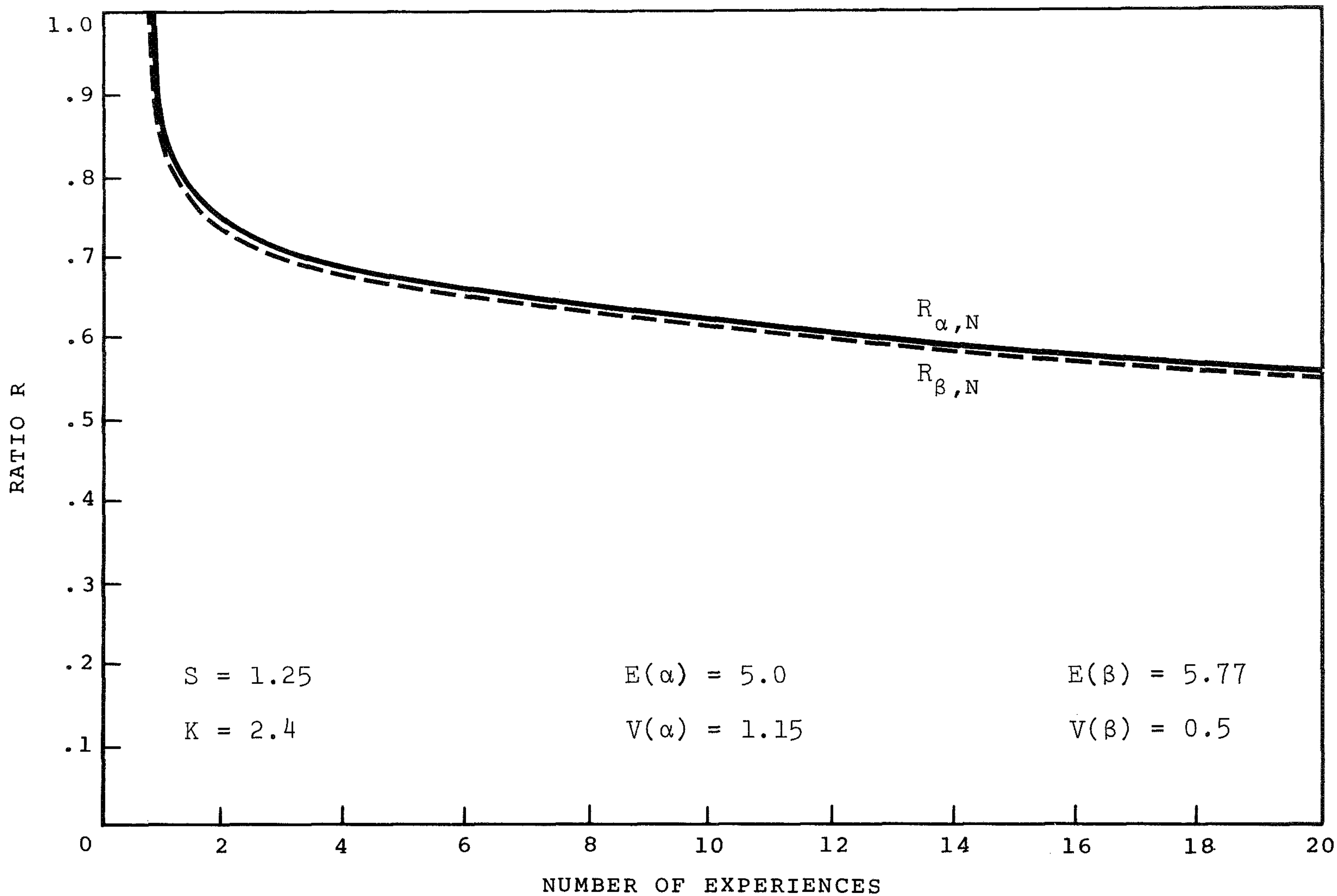

Figure 41. – $R_{\alpha,N}$ and $R_{\beta,N}$ for $Z = 2.0$; U-shaped prior.

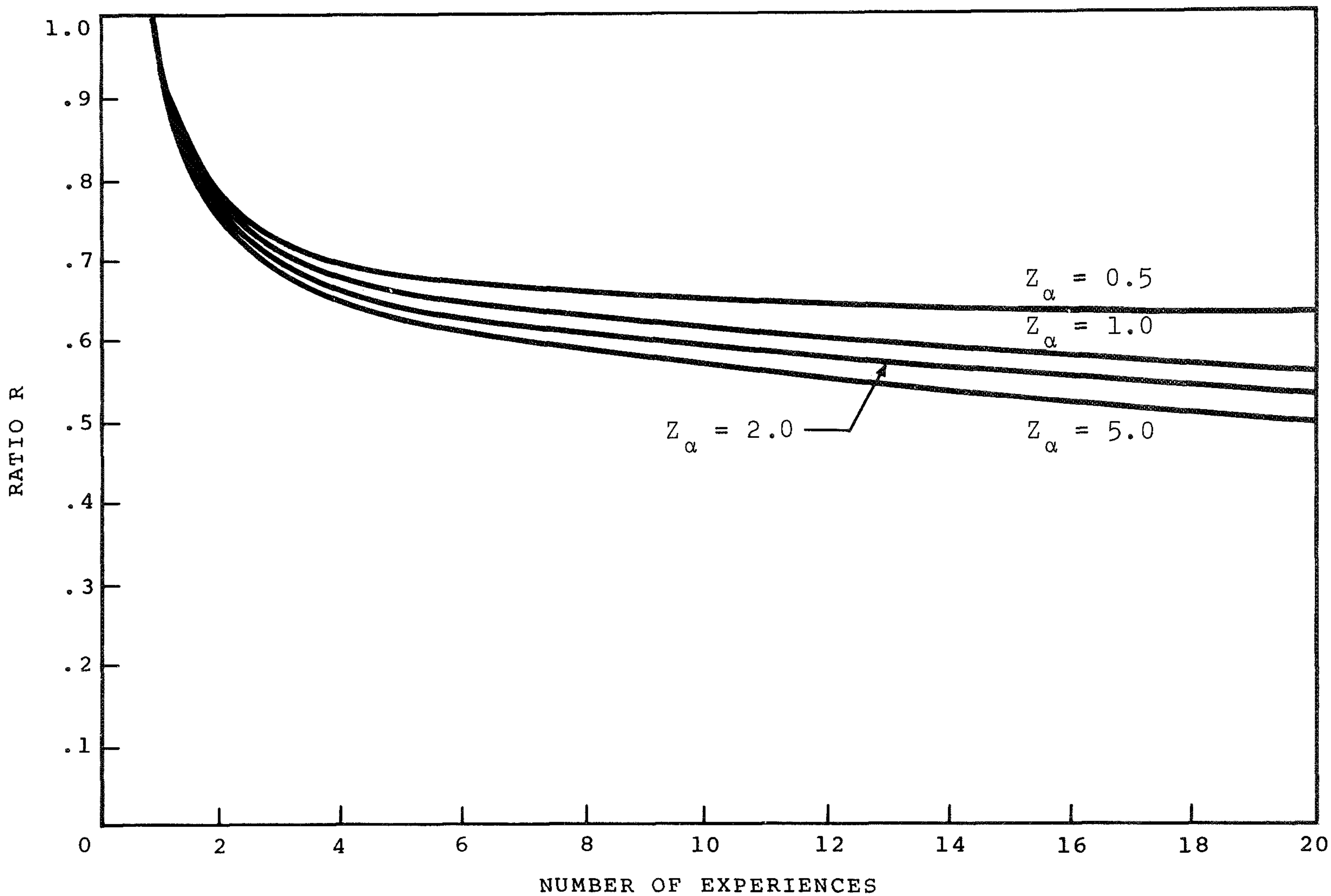

Figure 42. — $R_{\alpha,N}$ for several values of Z with $\rho = 0.0$.

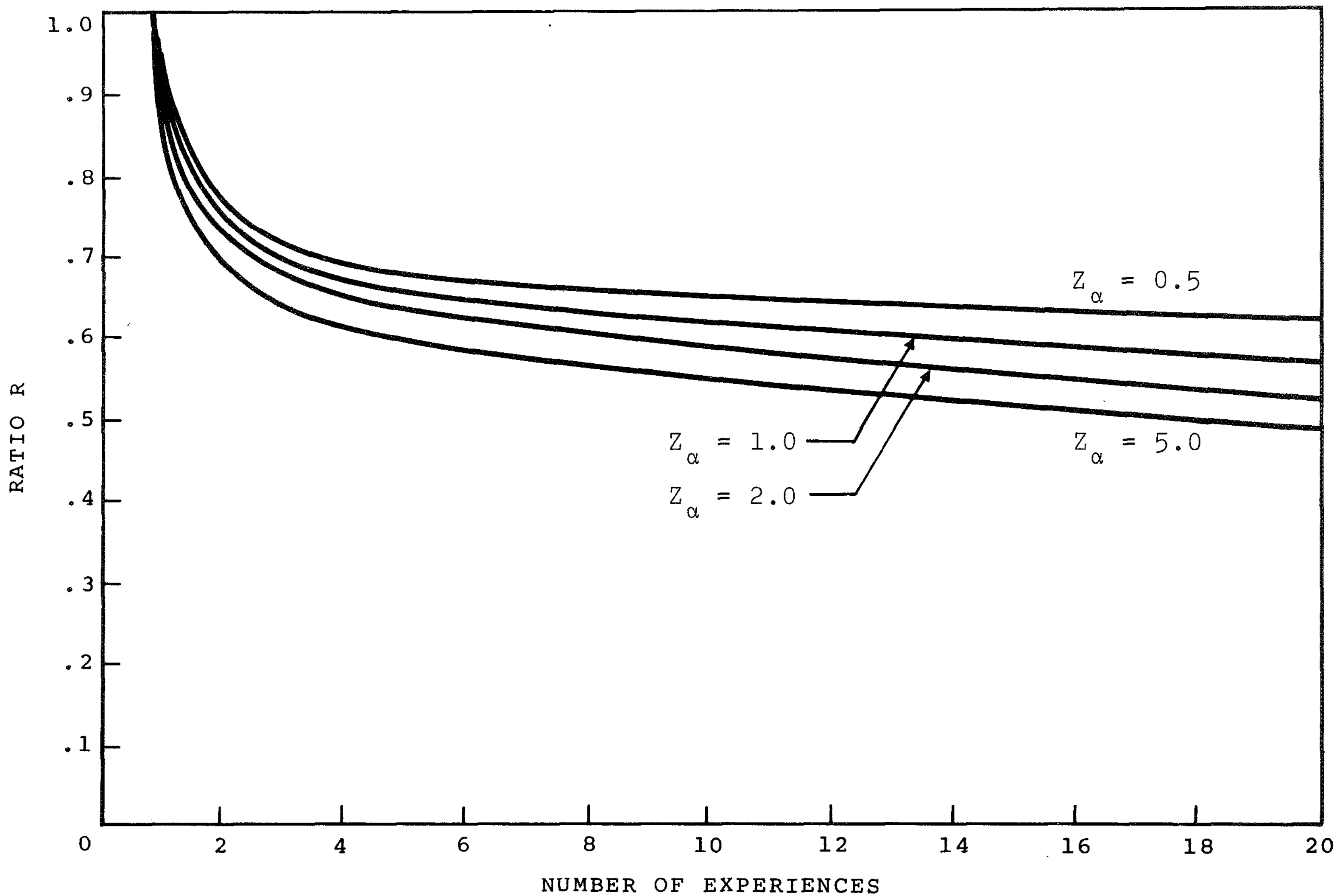

Figure 43. – $R_{\alpha,N}$ for several values of Z with $\rho = 0.5$.

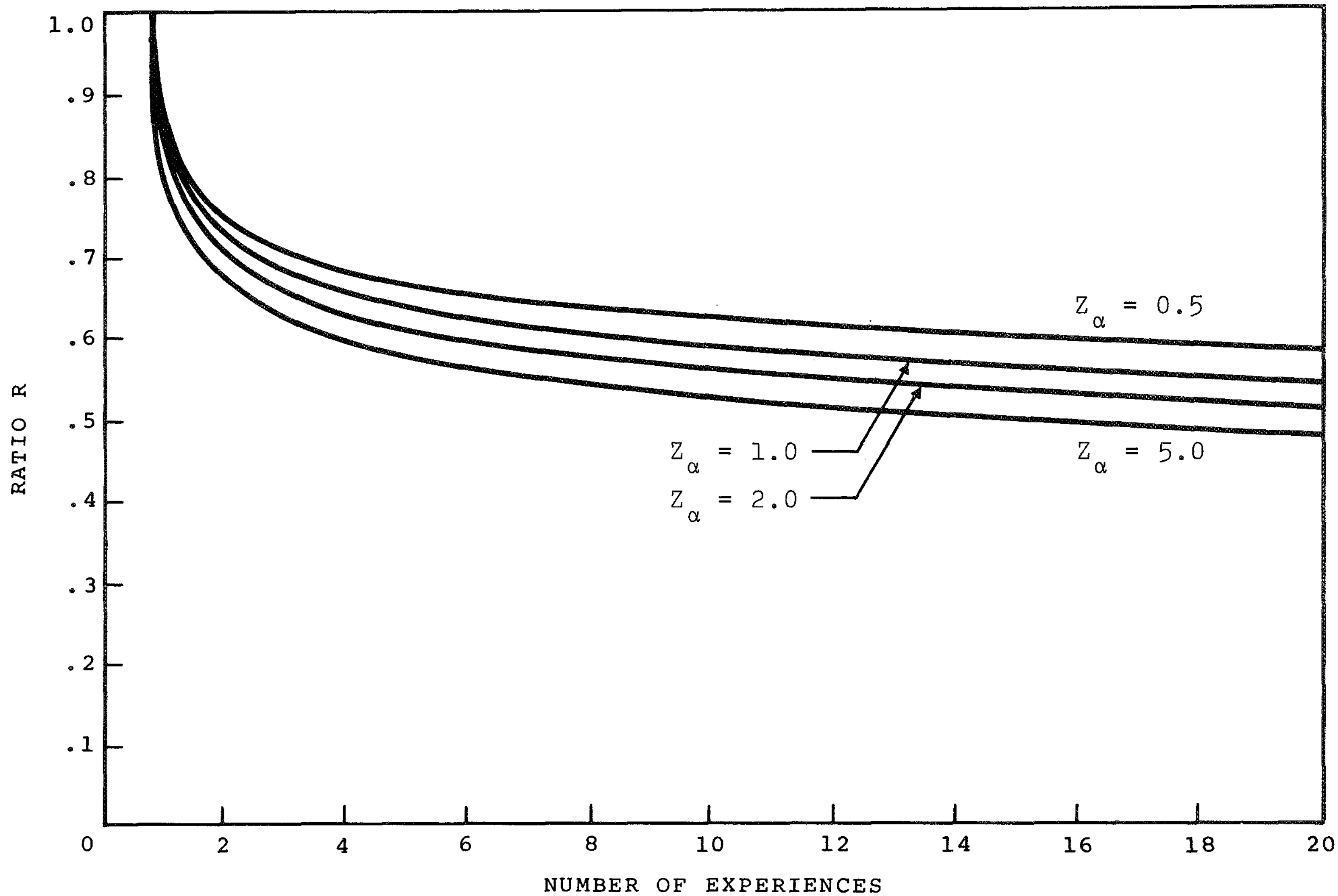

Figure 44. – $R_{\alpha,N}$ for several values of Z with $\rho = 0.9$.

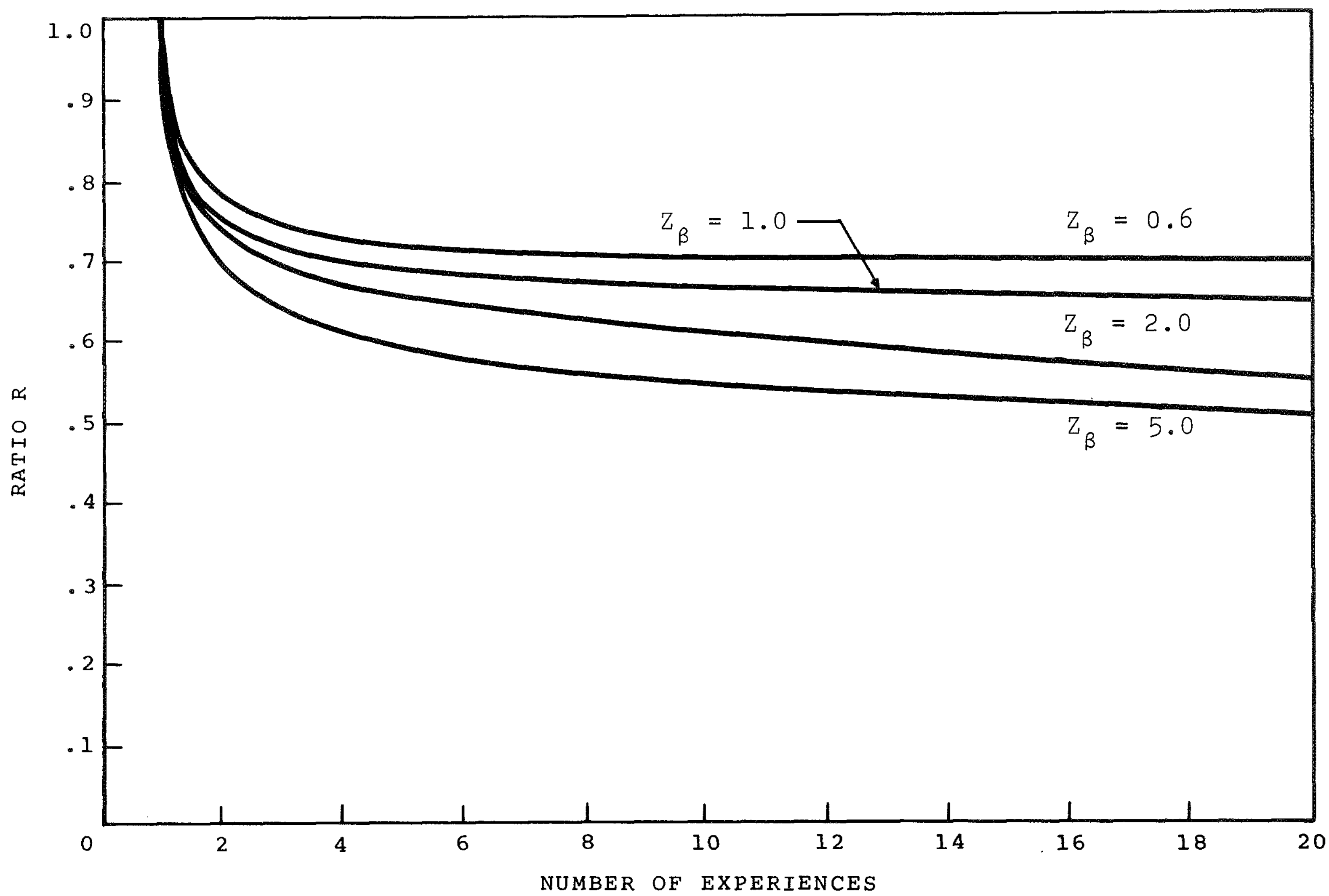

Figure 45. — $R_{\beta,N}$ for several values of Z with $\rho = 0.0$.

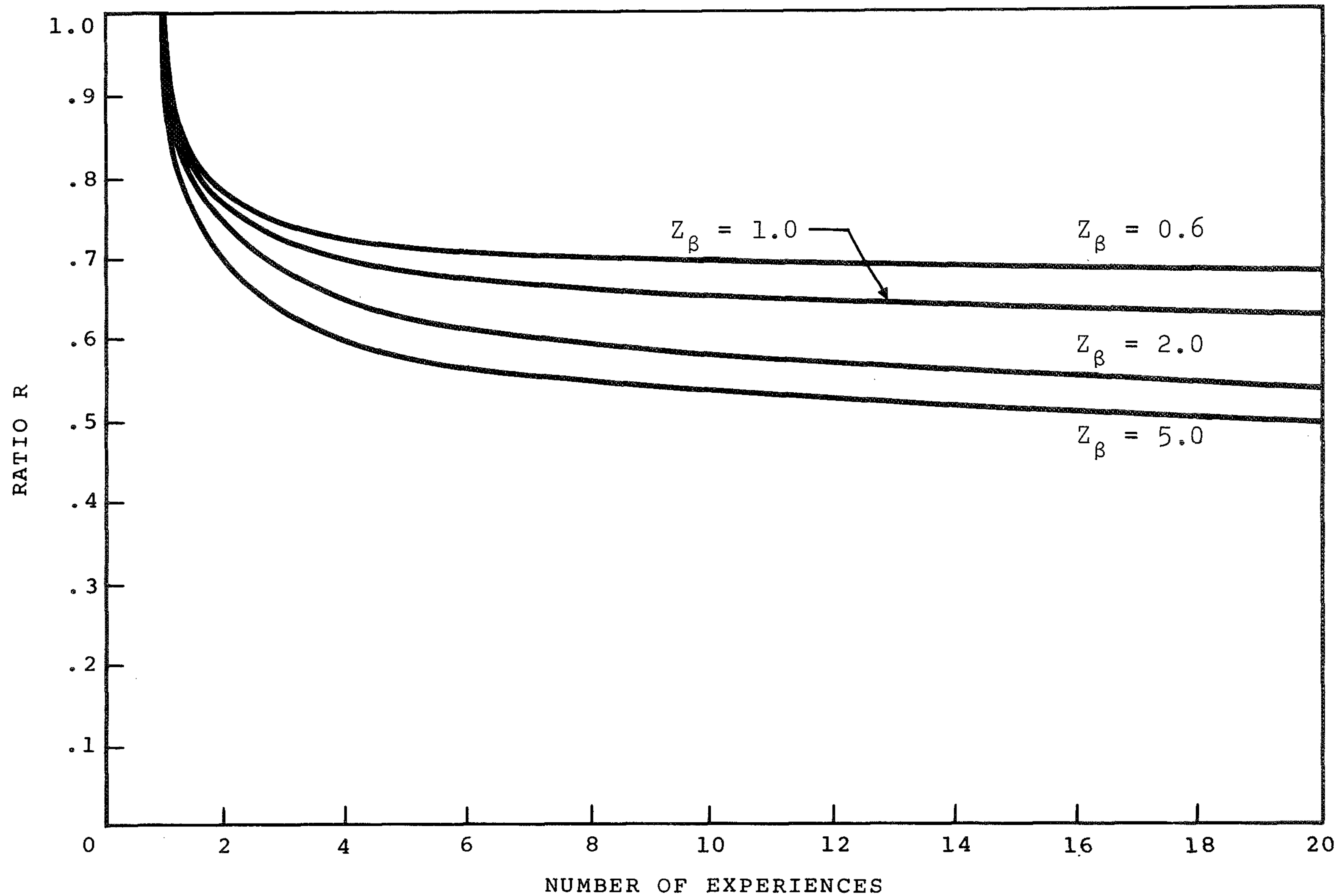

Figure 46. – $R_{\beta,N}$ for several values of Z with $\rho = 0.5$.

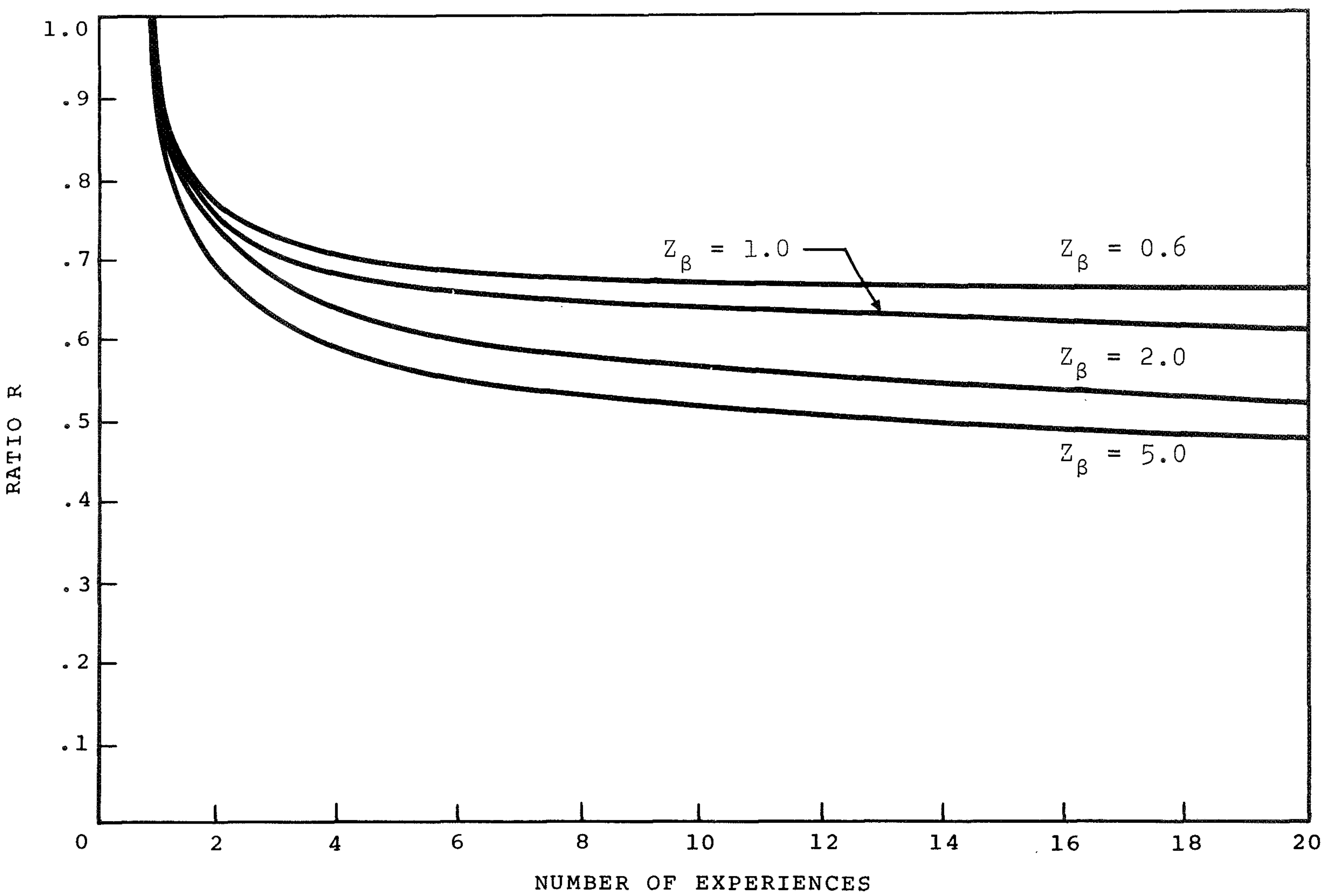

Figure 47. — $R_{\beta,N}$ for several values of Z with $\rho = 0.9$.

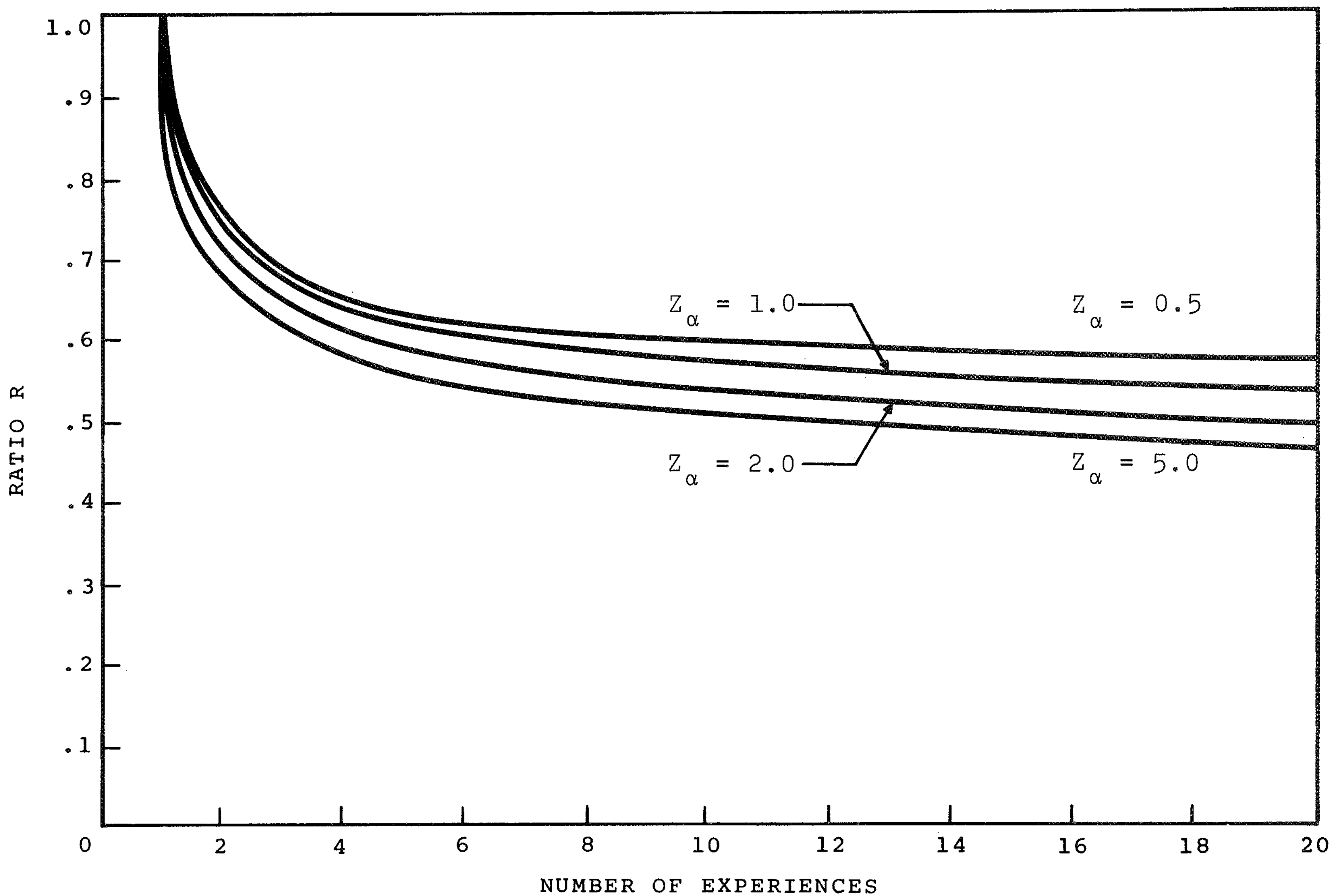

Figure 48. – $R_{\alpha,M}$ for several values of Z .

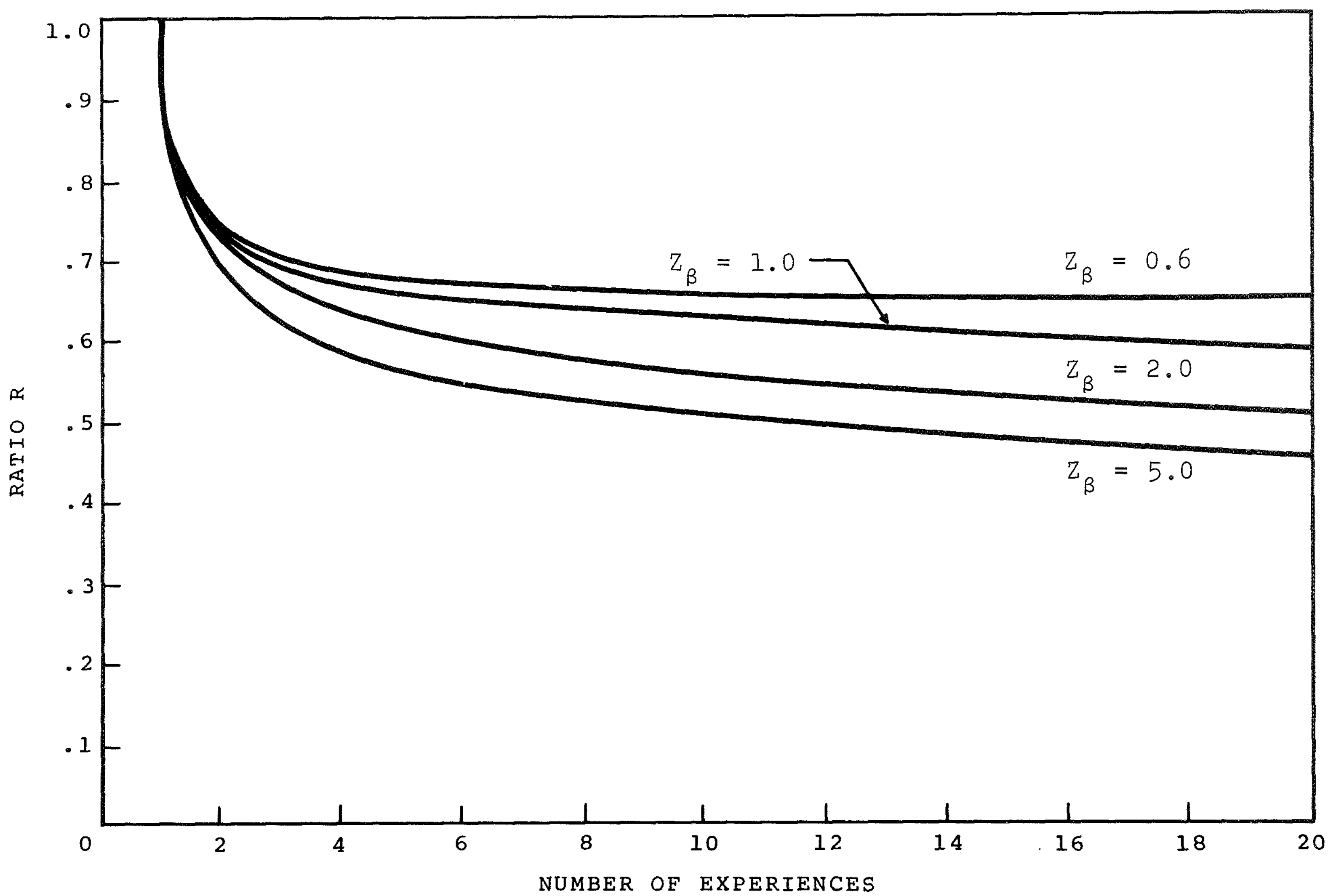

Figure 49. – $R_{\beta,M}$ for several values of Z .

CHAPTER VII

COMPARISONS WITH TWO ALTERNATIVE EMPIRICAL BAYES ESTIMATORS

In this chapter two alternative empirical Bayes estimators are considered. Where applicable, these estimators are applied to the distributions considered in the preceding chapters. Results from Monte Carlo simulation with these estimators are reported and directly compared with the results obtained from the corresponding continuously smooth empirical Bayes estimators.

7.1 Alternative Empirical Bayes Estimators

The methods of empirical Bayes estimation can be partitioned into two distinct classes. The first class consists of those methods which attempt to obtain empirical Bayes estimators without requiring explicit estimation of the prior distribution. The well-defined families developed by Rutherford and Krutchkoff [28] typify this technique. The second class consists of those methods which endeavor to obtain empirical Bayes estimators by considering an approximation to the prior distribution function. The method proposed by Lemon and Krutchkoff [17] demonstrates this technique. In a well-defined sense, the continuously smooth estimator does not belong to any of these classes. The continuously smooth estimator, however, can be considered an analog to the second method of estimation since it too attempts to approximate some form of the prior distribution.

Rutherford and Krutchkoff [28] established several well-defined families of distribution functions. Each family provides a unique empirical Bayes estimator $E_n(\theta|\underset{\sim}{x})$ having the property that

$$\underset{n\to\infty}{\text{p lim}}\ E_n(\theta|\underset{\sim}{x}) \quad = \quad E(\theta|\underset{\sim}{x}) \tag{7.1}$$

for all x . This property was considered desirable since previously [29] they had shown that ε-asymptotic optimality could be obtained by a truncated version of consistent estimators for the Bayes estimator. In practical situations, ε-asymptotic optimality is equivalent to asymptotic optimality which is defined by

$$\underset{n\to\infty}{\text{p lim}}\ R(\underset{\sim}{\delta}_n, G) \quad = \quad R(G) \quad . \tag{7.2}$$

Here $R(\cdot,\cdot)$ represents the overall risk, $\underset{\sim}{\delta}_n$ an empirical Bayes decision function, and $R(G)$ the Bayes risk. These elements were described in detail in section 1.4.

In particular we will be concerned with families F_1 and F_3 . For completeness we include their definitions. A family of distributions $\{F(x|\theta):\theta\varepsilon\Theta\}$ is said to be a member of F_1 if

i) the random variable X is discrete for each θ, and

ii) the probability mass function $P(x|\theta)$ is such that

$$\frac{P(x+1|\theta)}{P(x|\theta)} = a(x) + \theta b(x)$$

where $a(x)$ and $b(x)$ are any functions such that $b(x) \neq 0$. If $P_n(x)$ and $P_n(x+1)$ are consistent estimators for the marginal probability mass function $P(x)$, then a consistent empirical Bayes estimate of θ_n is given by

$$\theta_D^* = \frac{P_n(x_n+1)}{b(x_n)\,P(x_n)} - \frac{a(x_n)}{b(x_n)} .$$

A family of distributions $\{F(x|\theta):\theta\varepsilon\Theta\}$ is said to be a member of F_2 if

i) X is a continuous random variable for all $\theta\varepsilon\Theta$, and

ii) the probability densities $f(x|\theta)$ are such that

$$\frac{\frac{\partial f(x|\theta)}{\partial x}}{f(x|\theta)} = a(x) + \theta b(x)$$

where a(x) and b(x) are any functions such that $b(x) \neq 0$. If $f_n(x)$ and $f_n'(x)$ are consistent estimators for the marginal density and its first derivative respectively, then a consistent empirical Bayes estimate of θ_n is given by

$$\theta_D^* = \frac{f_n'(x_n)}{b(x_n) f_n(x_n)} - \frac{a(x_n)}{b(x_n)} \; .$$

For a discussion on estimating $f_n'(x)$, see Rutherford [27].

Lemon and Krutchkoff [17] proposed a general smoothing technique for obtaining empirical Bayes estimators. One particular estimator is essentially obtained by the replacement of the prior distribution by a step function having steps of equal height 1/n at each of n previous classical estimates. This estimator can be represented by

$$\hat{\hat{\theta}}_D = \frac{\sum_{i=1}^{n} \hat{\theta}_i f(\hat{\theta}_n | \hat{\theta}_i)}{\sum_{i=1}^{n} f(\hat{\theta}_n | \hat{\theta}_i)} \tag{7.3}$$

where $\hat{\theta}_i$ represents the *ith* classical estimate of θ and $f(\cdot|\cdot)$ is the kernel of integration.

Also suggested was a second iteration of $\hat{\hat{\theta}}_D$. This iteration gives

$$\hat{\hat{\theta}}_D' = \frac{\sum_{i=1}^{n} \hat{\hat{\theta}}_{D,i} f(\hat{\theta}_n | \hat{\hat{\theta}}_{D,i})}{\sum_{i=1}^{n} f(\hat{\theta}_n | \hat{\hat{\theta}}_{D,i})} \tag{7.4}$$

where $\hat{\hat{\theta}}_{D,i}$ is given by (7.3). Further iterations are, of course, possible; however, as with the continuously smooth estimators, they may provide a significant increase in squared error and thus may be undesirable.

7.2 The Poisson Distribution

Let us assume that in each experiment, a single observation x_i $(i = 1,2,\cdots,n)$ is obtained from the Poisson mass function given by (3.1). From (3.4) the maximum-likelihood estimate of θ is simply x or $\hat{\theta}$. The subscript k has, for convenience, been deleted.

The Poisson distribution is readily verified to be a member of family F_1 in which $a(\hat{\theta}_n) = 0$ and $b(\hat{\theta}_n) = 1/(\hat{\theta}_n + 1)$. The empirical Bayes estimate of θ_n is therefore given by

$$\theta_P^* = (\hat{\theta}_n + 1) \frac{P_n(\hat{\theta}_n + 1)}{P_n(\hat{\theta}_n)} \quad . \tag{7.5}$$

Clearly, $P_n(\hat{\theta}_n)$ can be estimated from the sequence of observations $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_n$. Although θ_P^* is the unique estimator obtained from family F_1 , it is precisely the estimator Robbins [26] used to introduce the empirical Bayes situation.

For the Poisson distribution, the estimators proposed by Lemon and Krutchkoff [17] can be represented by

$$\hat{\hat{\theta}}_P = \frac{\sum_{i=1}^{n} e^{-\hat{\theta}_i} \hat{\theta}_i^{\hat{\theta}_n + 1}}{\sum_{i=1}^{n} e^{-\hat{\theta}_i} \hat{\theta}_i^{\hat{\theta}_n}} \tag{7.6}$$

and

$$\hat{\hat{\theta}}'_P = \frac{\sum_{i=1}^{n} e^{-\hat{\hat{\theta}}_{P,i}} \hat{\hat{\theta}}_{P,i}^{\hat{\theta}_n + 1}}{\sum_{i=1}^{n} e^{-\hat{\hat{\theta}}_{P,i}} \hat{\hat{\theta}}_{P,i}^{\hat{\theta}_n}} \tag{7.7}$$

respectively.

Monte Carlo simulations for the estimators θ_P^* , $\hat{\hat{\theta}}_P$, and $\hat{\hat{\theta}}'_P$ were conducted in a manner analogous to the procedure described for $\tilde{\theta}_D$ in section 3.5. As in section 3.5 the criterion for comparison was mean-squared error, and therefore the ratio R defined by (3.25) was of interest. Here we denote this ratio by

R_P^* , $\hat{R}_P$, and $\hat{R}_P'$ when calculated with Θ_P^* , $\hat{\Theta}_P$, and $\hat{\Theta}_P'$ respectively. The ratio R calculated with the smooth estimator $\tilde{\Theta}_{P,V}$ given by (3.22) will be denoted by $R_{P,V}$.

In Figures 50-52 the values of $\hat{R}_P'$ and $R_{P,V}$, represented by the broken and solid lines respectively, are plotted as a function of the number of experiences. These plots are given for various values of the summary quantity Z given by (3.27). The estimator $\hat{\Theta}_P'$ was observed to provide uniform squared-error improvement over the estimator $\hat{\Theta}_P$; therefore the ratio $\hat{R}_P$ is not shown. Also, the ratio R_P^* is omitted since for $0.5 \leq Z \leq 5.0$ and $n \leq 20$, the maximum-likelihood estimator gave smaller mean-squared errors than the estimator Θ_P^* regardless of the form of the prior distribution. By comparing the values of $R_{P,V}$ with $\hat{R}_P'$ in each of the figures, it can be seen that the continuously smooth estimator provides consistent mean-squared improvement over the estimator $\hat{\Theta}_P'$. Hence, $\tilde{\Theta}_{P,V}$ provides consistent improvement over $\hat{\Theta}_P$ and, of course, over Θ_P^* .

7.3 The Weibull Distribution with Known Shape Parameter

Let us assume that in each experiment, a random sample of k observations $\underset{\sim}{x} = (x_1, x_2, \cdots, x_k)$ is obtained from the Weibull density function given by (4.1). In section 4.1, the sufficient statistic $T = \sum_{j=1}^{k} x_j^{\beta}$ formed with this sample was shown to have the conditional gamma density function given by (4.15). Differentiating this function $f(t|\alpha)$ with respect to T and dividing by $f(t|\alpha)$ we obtain

$$\frac{\frac{\partial f(t|\alpha)}{\partial t}}{f(t|\alpha)} = -\alpha + \frac{k-1}{t} \quad . \tag{7.8}$$

The gamma distribution is therefore a member of family F_2, and a consistent estimator for α_n can be given by

$$\alpha_G^* = \frac{k-1}{t_n} - \frac{f_n'(t_n)}{f_n(t_n)} \tag{7.9}$$

The consistent estimators $f_n(t_n)$ and $f_n'(t_n)$ chosen to represent $f(t)$ and its first derivative are given by

$$f_n(t_n) = \frac{1}{2\pi nh} \sum_{i=1}^{n} \left[\frac{\sin\left(\frac{t_n - t_i}{2h}\right)}{\left(\frac{t_n - t_i}{2h}\right)} \right]^2 \tag{7.10}$$

and

$$f'(t_n) = \frac{f_n(t_n + h) - f_n(t_n)}{h} \tag{7.11}$$

respectively.

The estimators of Lemon and Krutchkoff [17] $\hat{\hat{\alpha}}_G$ and $\hat{\hat{\alpha}}'_G$ can be constructed using as the kernel of integration the inverted gamma density function given by (4.16). These estimators become

$$\hat{\hat{\alpha}}_G = \frac{\left(\frac{\hat{\alpha}_i k}{\hat{\alpha}_n}\right)^{k+1} e^{-\left(\frac{\hat{\alpha}_i k}{\hat{\alpha}_n}\right)}}{\frac{1}{\hat{\alpha}_i}\left(\frac{\hat{\alpha}_i k}{\hat{\alpha}_n}\right)^{k+1} e^{-\left(\frac{\hat{\alpha}_i k}{\hat{\alpha}_n}\right)}} \tag{7.12}$$

and

$$\hat{\hat{\alpha}}'_G = \frac{\left(\frac{\hat{\hat{\alpha}}_{G,i} k}{\hat{\alpha}_n}\right)^{k+1} e^{-\left(\frac{\hat{\hat{\alpha}}_{G,i} k}{\hat{\alpha}_n}\right)}}{\frac{1}{\hat{\hat{\alpha}}_{G,i}}\left(\frac{\hat{\hat{\alpha}}_{G,i} k}{\hat{\alpha}_n}\right)^{k+1} e^{-\left(\frac{\hat{\hat{\alpha}}_{G,i} k}{\hat{\alpha}_n}\right)}} \tag{7.13}$$

where $\hat{\alpha}_i$ $(i = 1,2,\cdots,n)$ is the maximum-likelihood estimate from the *ith* experiment.

Monte Carlo simulations for the empirical Bayes estimators α_G^* , $\hat{\hat{\alpha}}_G$, and $\hat{\hat{\alpha}}_G'$ were conducted in a manner analogous to the procedure described for $\tilde{\theta}_D$ in section 3.5. The ratios R_G^* , $\hat{\hat{R}}_G$, and $\hat{\hat{R}}_G'$ defined by (3.25) were calculated with α_G^* , $\hat{\hat{\alpha}}_G$, and $\hat{\hat{\alpha}}_G'$ respectively.

In Figures 53-55, the values of R_G^* , $\hat{\hat{R}}_G'$, and $R_{G,V}$, denoted by the dotted, broken, and solid lines respectively, are plotted for various values of Z defined by (4.44). The ratio $R_{G,V}$ was calculated with the continuously smooth estimator $\tilde{\alpha}_{G,V}$ given by (4.42). The ratio $\hat{\hat{R}}_G$ has been omitted in each figure since the results obtained with the estimator $\hat{\hat{\alpha}}_G'$ were in all cases considered uniformly superior when compared to the results obtained with $\hat{\hat{\alpha}}_G$. In each figure comparison of the plots of R_G^* , $\hat{\hat{R}}_G'$, and $R_{G,V}$ as a function of the number of experiments demonstrates the significant improvements one can obtain using the continuously smooth estimator $\tilde{\alpha}_{G,V}$ as opposed to $\hat{\hat{\alpha}}_G'$ or α_G^* . Since $\tilde{\alpha}_{G,V}$ provides consistent mean-squared improvement over $\hat{\hat{\alpha}}_G'$, it also provides consistent improvement over $\hat{\hat{\alpha}}_G$.

7.4 The Weibull Distribution with Known Scale Parameter

Let us assume that in each experiment a random sample of k observations is obtained from the Weibull density function given by (5.1). The maximum-likelihood estimator for the shape parameter β is found by the solution of (5.3). The distribution of this estimator, given any value of the parameter β, has an asymptotic normal distribution with mean β and variance given by (5.12). This distribution can be used to form the empirical Bayes estimators

$$\hat{\hat{\beta}}_N = \frac{\displaystyle\sum_{i=1}^{n} e^{-\frac{k}{2}\left(\frac{\hat{\beta}_n-\hat{\beta}_i}{\hat{\beta}_i}\right)^2}}{\displaystyle\sum_{i=1}^{n} \frac{1}{\hat{\beta}_i}\, e^{-\frac{k}{2}\left(\frac{\hat{\beta}_n-\hat{\beta}_i}{\hat{\beta}_i}\right)^2}} \tag{7.14}$$

and

$$\hat{\hat{\beta}}{}'_N = \frac{\displaystyle\sum_{i=1}^{n} e^{-\frac{k}{2}\left(\frac{\hat{\beta}_n-\hat{\hat{\beta}}_{N,i}}{\hat{\hat{\beta}}_{N,i}}\right)^2}}{\displaystyle\sum_{i=1}^{n} \frac{1}{\hat{\hat{\beta}}_{N,i}}\, e^{-\frac{k}{2}\left(\frac{\hat{\beta}_n-\hat{\hat{\beta}}_{N,i}}{\hat{\hat{\beta}}_{N,i}}\right)^2}} \tag{7.15}$$

The estimators given by (7.14) and (7.15) are of the form proposed by Lemon and Krutchkoff [17]; however, they are based on an asymptotic distribution rather than on the true distribution. This result represents a natural extension of their estimators and will be treated accordingly. We remark here that the Weibull distribution with known scale parameter cannot be placed into any of the families of Rutherford and Krutchkoff [28].

Monte Carlo simulations for the estimators $\hat{\hat{\beta}}_N$ and $\hat{\hat{\beta}}'_N$ were conducted in a manner analogous to the procedure described for $\tilde{\theta}_D$ in section 3.5. The ratios $\hat{R}_N$ and $\hat{\hat{R}}'_N$ defined by (3.25) were calculated with $\hat{\hat{\beta}}_N$ and $\hat{\hat{\beta}}'_N$ respectively.

In Figures 56-58 the values of $\hat{\hat{R}}'_N$ and $R_{N,V}$ denoted by the broken and solid lines respectively are plotted for various values of Z as defined by (5.25). The ratio $R_{N,V}$ was calculated with the continuously smooth estimator $\tilde{\beta}_{N,V}$ given by (5.23). As in the above cases, it was found that the estimator $\hat{\hat{\beta}}'_N$ provided significant squared-error improvement over the corresponding estimator $\hat{\hat{\beta}}_N$, and therefore the ratio $\hat{\hat{R}}_N$ has been omitted in each figure. Comparison of the plots of $\hat{\hat{R}}'_N$ and $R_{N,V}$ in each of the

Figures 56-58 demonstrates the significant improvement one can obtain using the continuously smooth estimator $\tilde{\beta}_{N,V}$ as opposed to $\hat{\hat{\beta}}'_N$. Since $\hat{\hat{\beta}}'_N$ was observed to give uniform improvement when compared with $\hat{\hat{\beta}}_N$, the estimator $\tilde{\beta}_{N,V}$ is also more efficient than $\hat{\hat{\beta}}_N$.

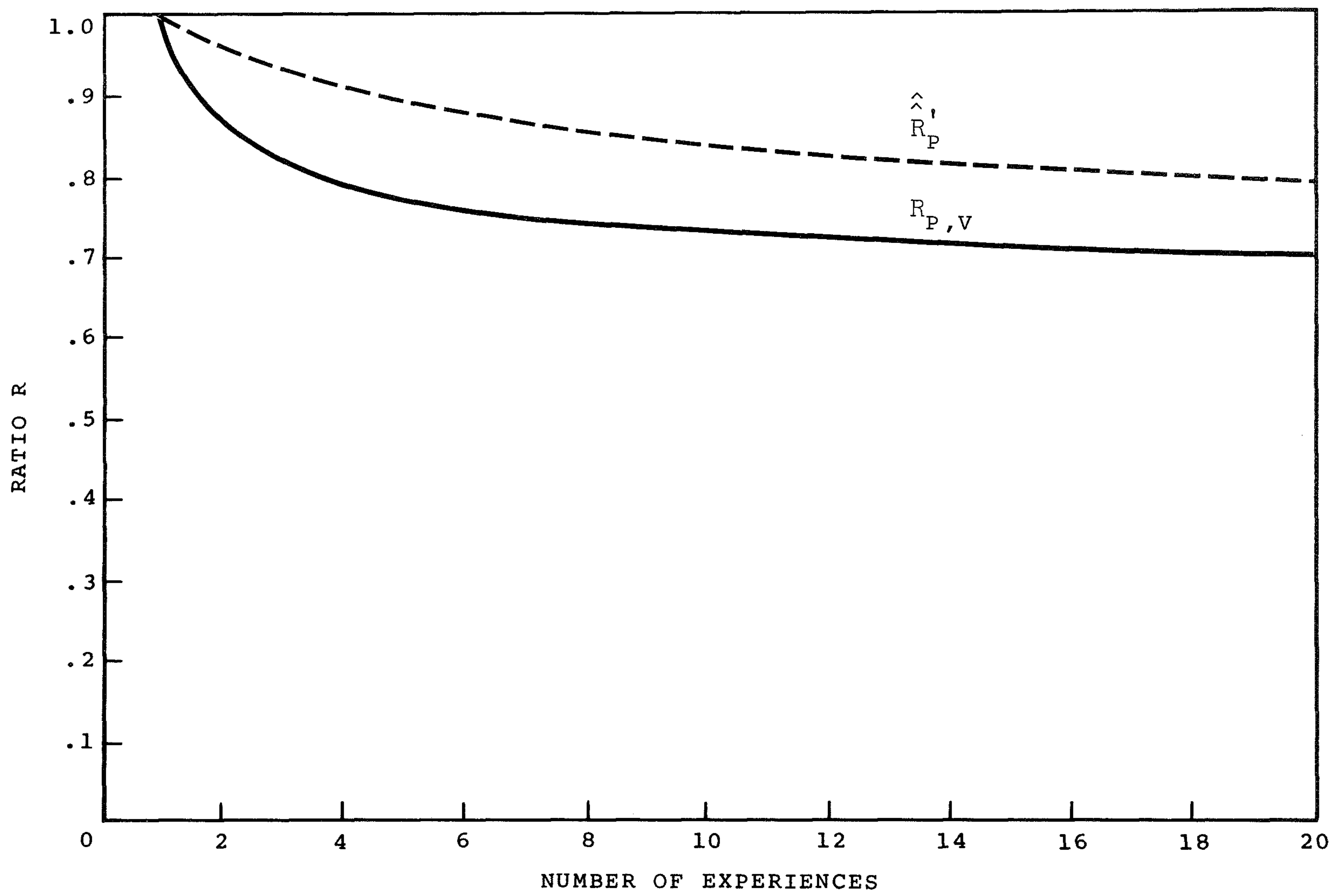

Figure 50. – $\hat{\hat{R}}_P'$ vs $R_{P,V}$ for $Z = 0.5$.

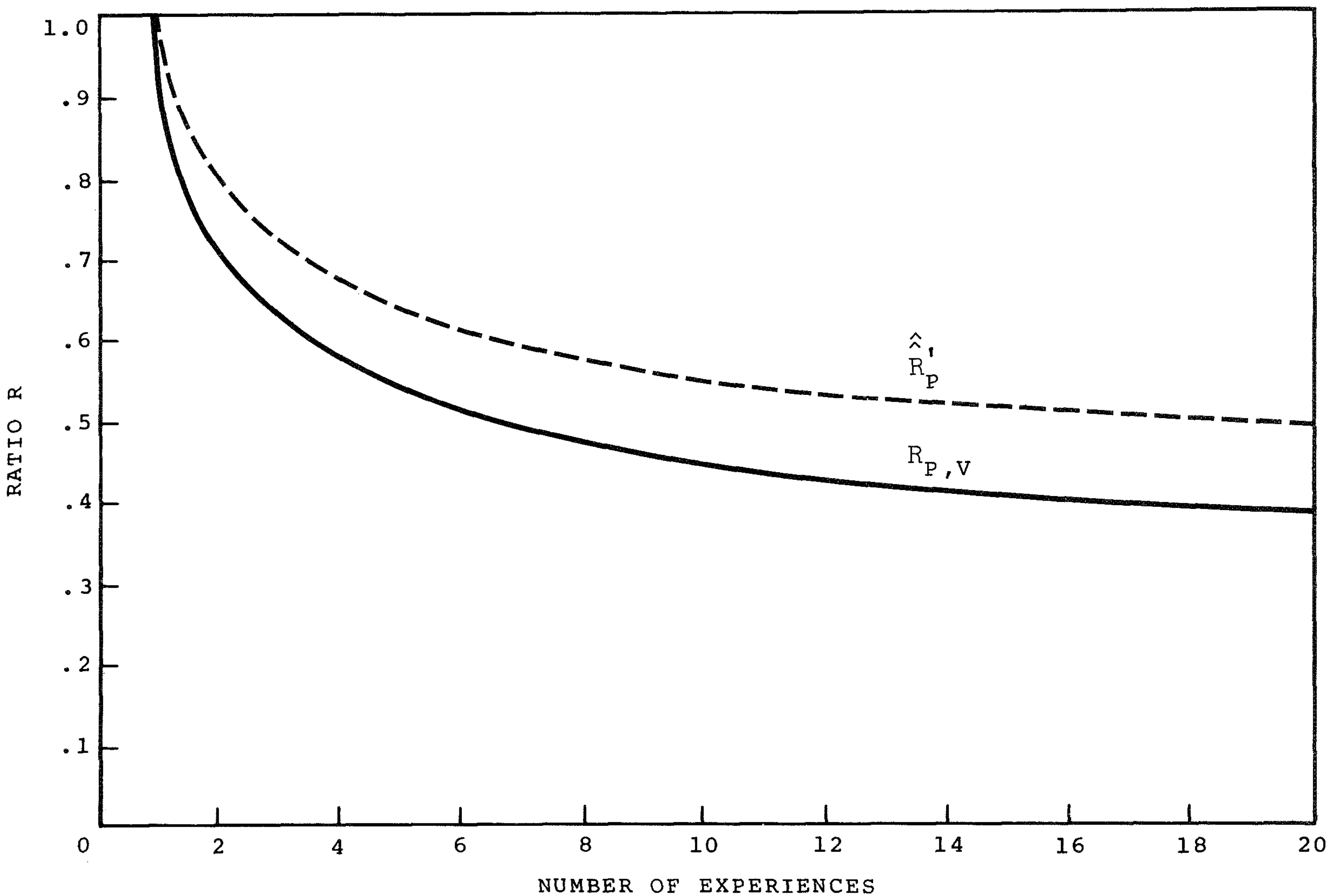

Figure 51. – $\hat{\hat{R}}'_P$ vs $R_{P,V}$ for Z = 2.0 .

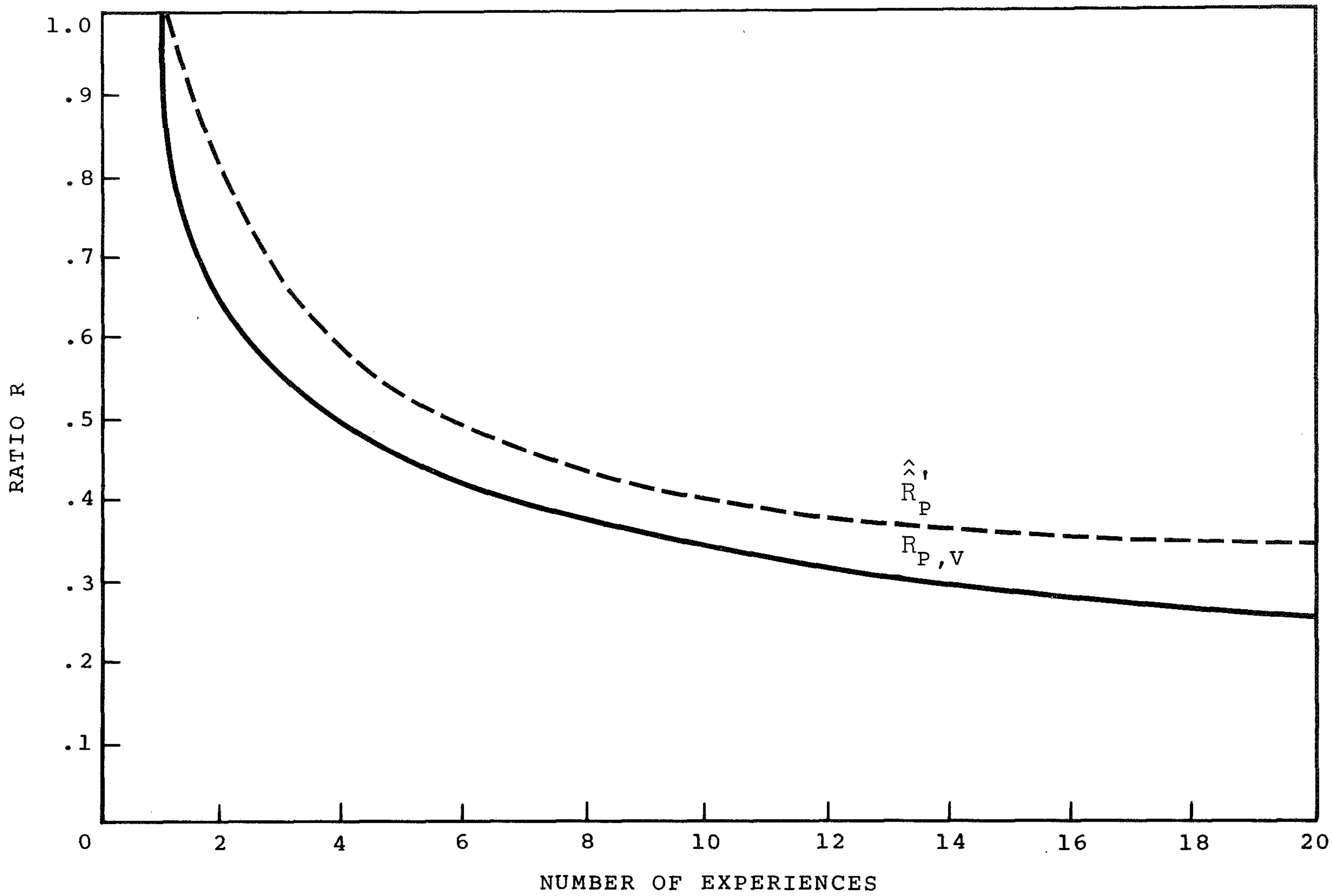

Figure 52. — $\hat{\hat{R}}_P'$ vs $R_{P,V}$ for $Z = 5.0$.

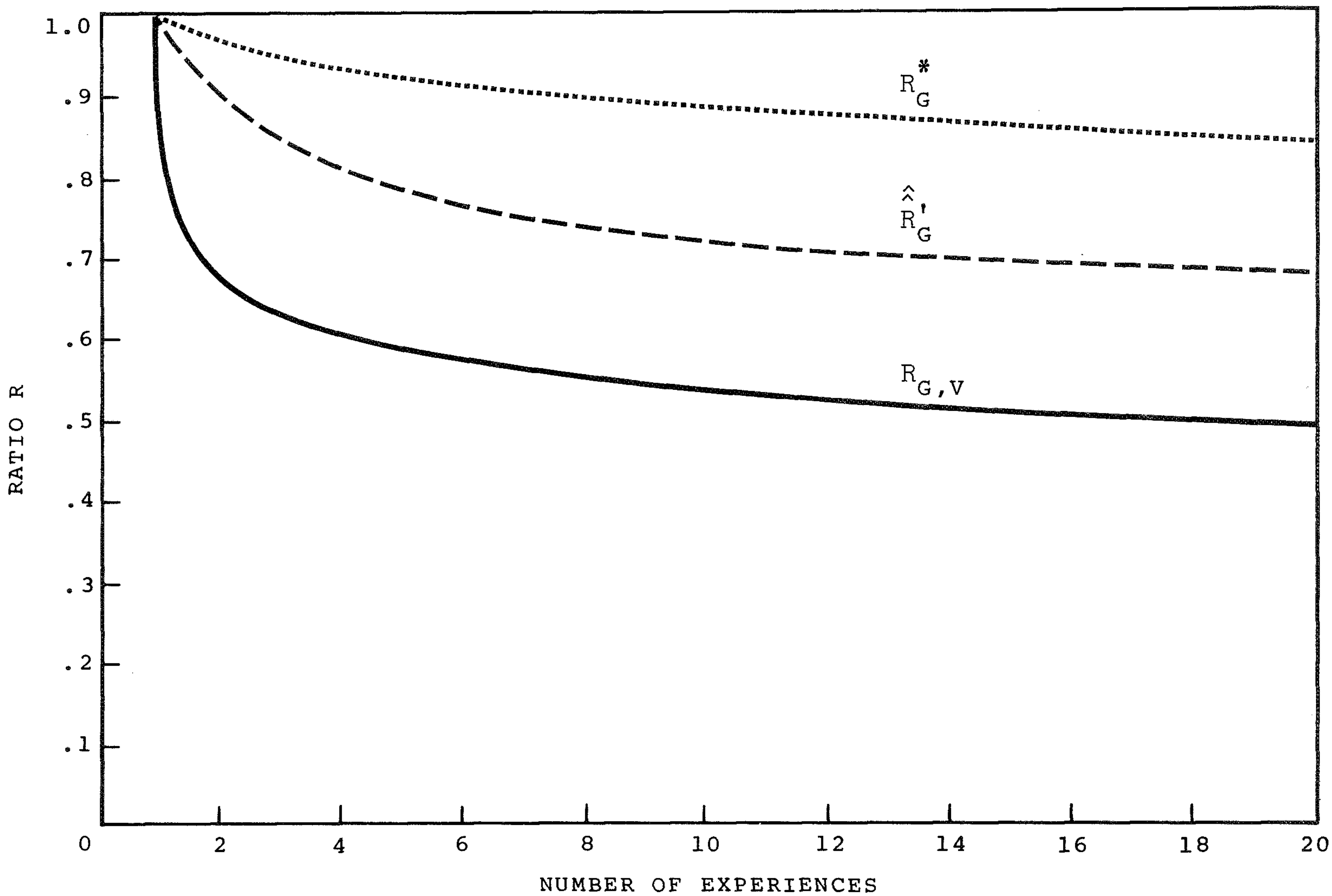

Figure 53. – R_G^* and $\hat{R}_G'$ vs $R_{G,V}$ for $Z = 0.5$.

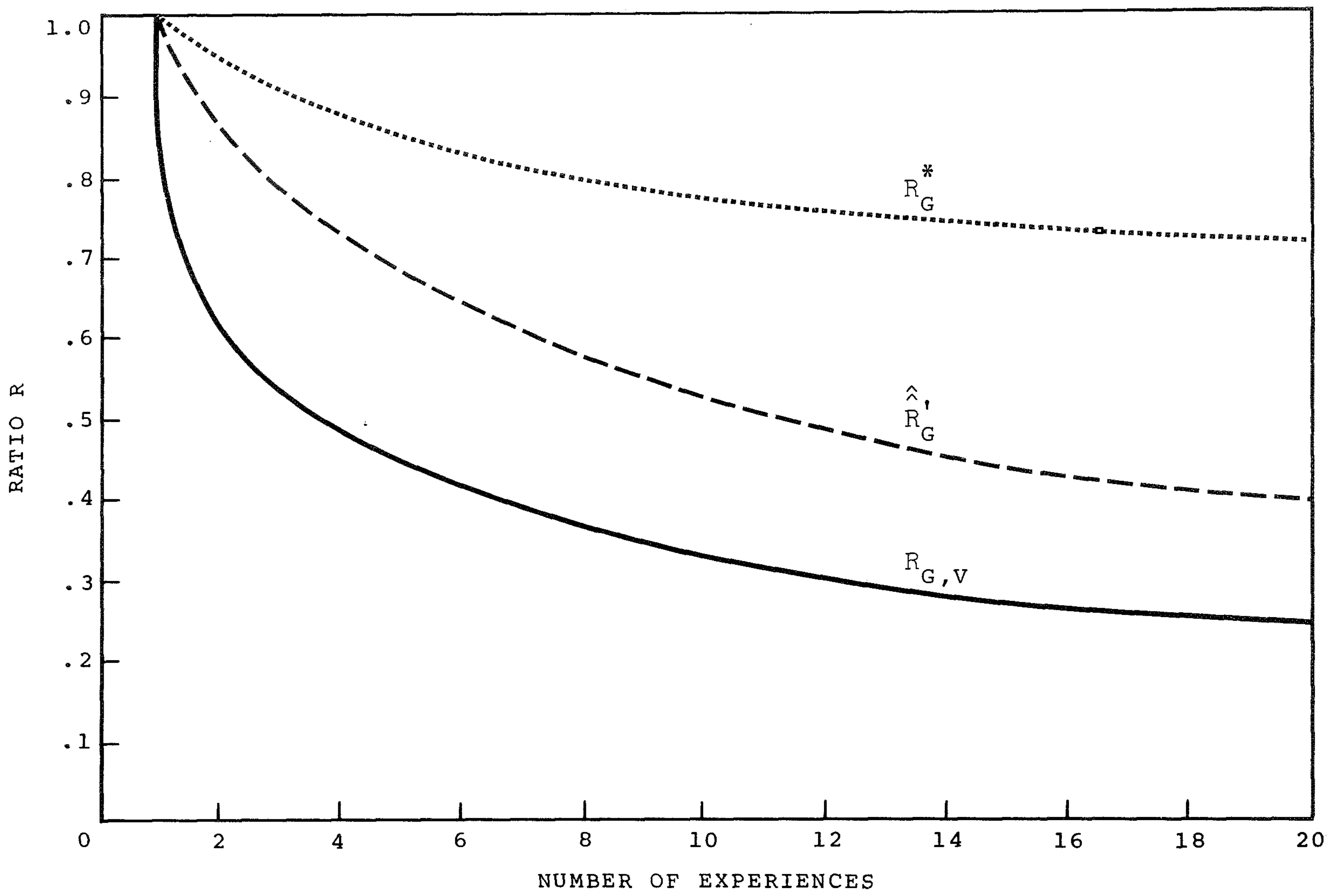

Figure 54. – R_G^* and $\hat{\hat{R}}_G'$ vs $R_{G,V}$ for $Z = 2.0$.

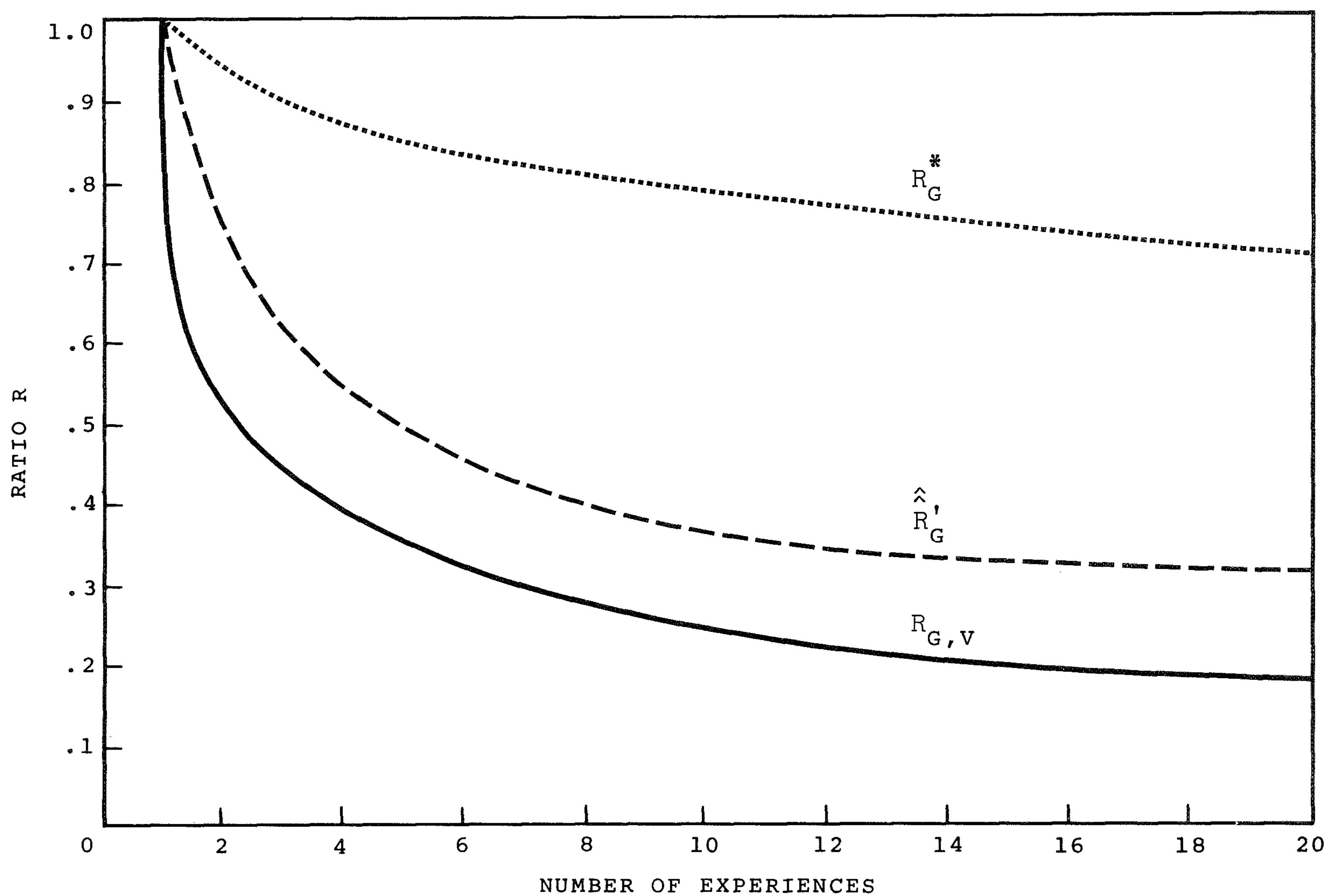

Figure 55. – R_G^* and $\hat{\hat{R}}'_G$ vs $R_{G,V}$ for $Z = 5.0$.

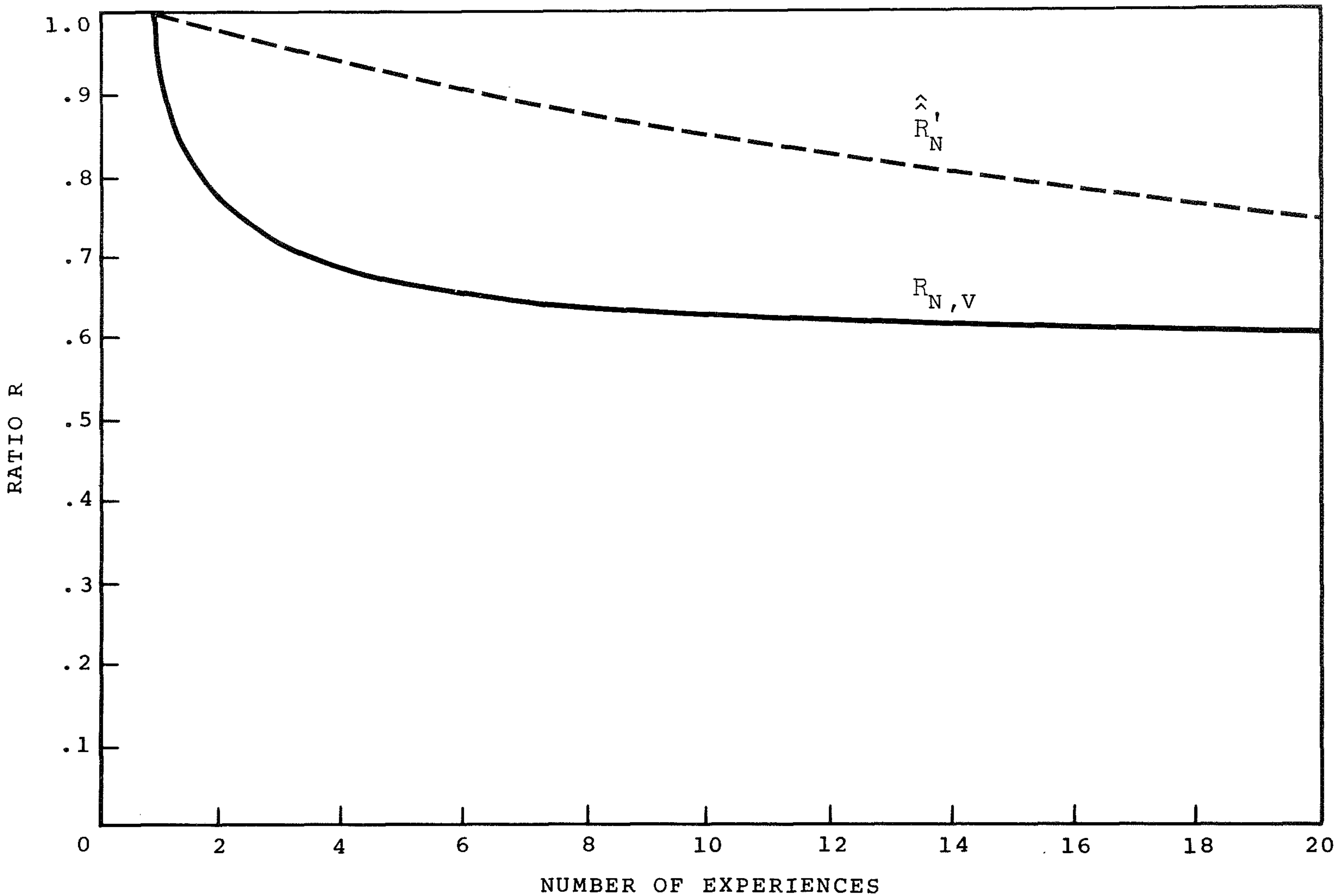

Figure 56. – $\hat{\hat{R}}'_N$ vs $R_{N,V}$ for $Z = 0.5$.

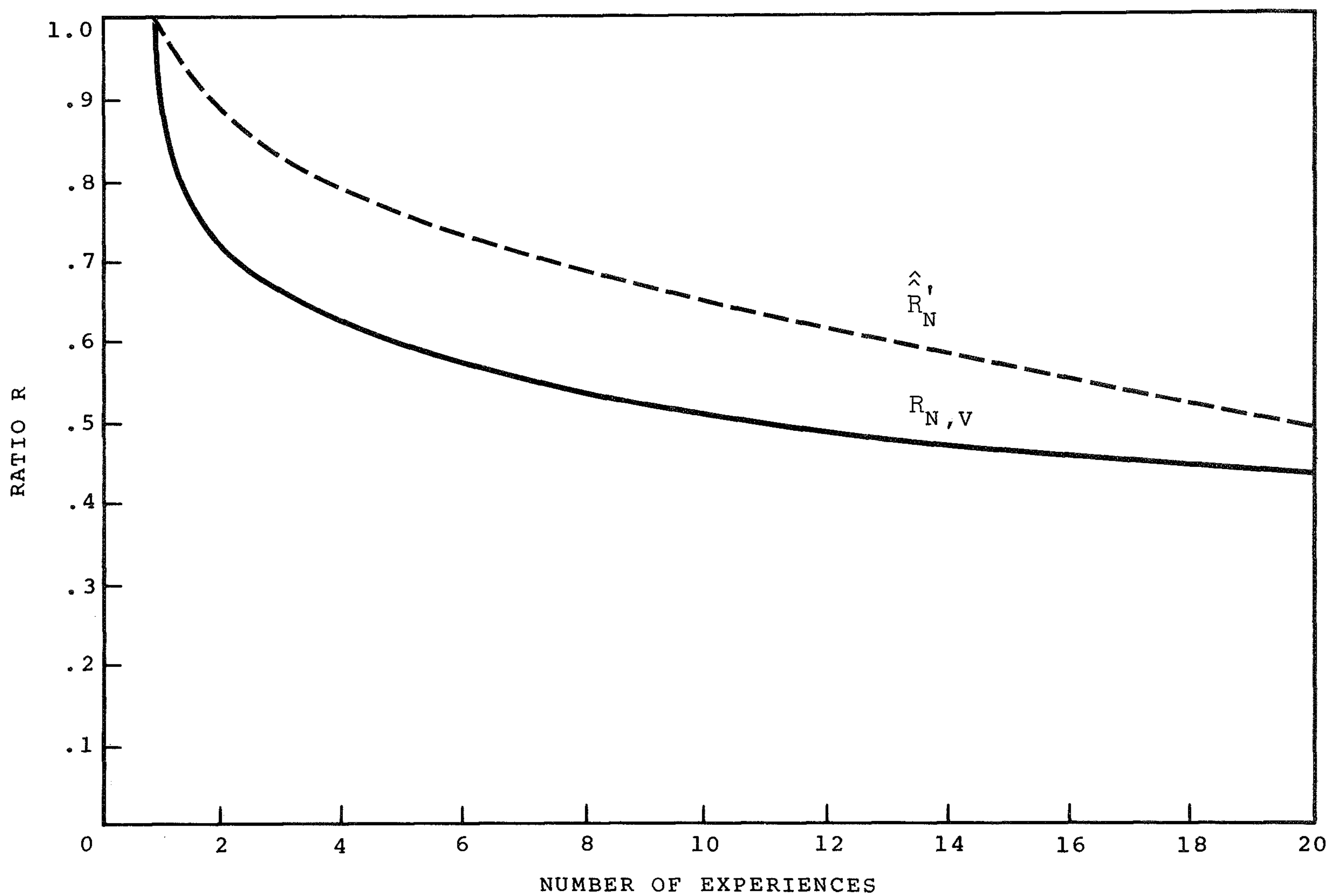

Figure 57. – $\hat{\hat{R}}'_N$ vs $R_{N,V}$ for $Z = 2.0$.

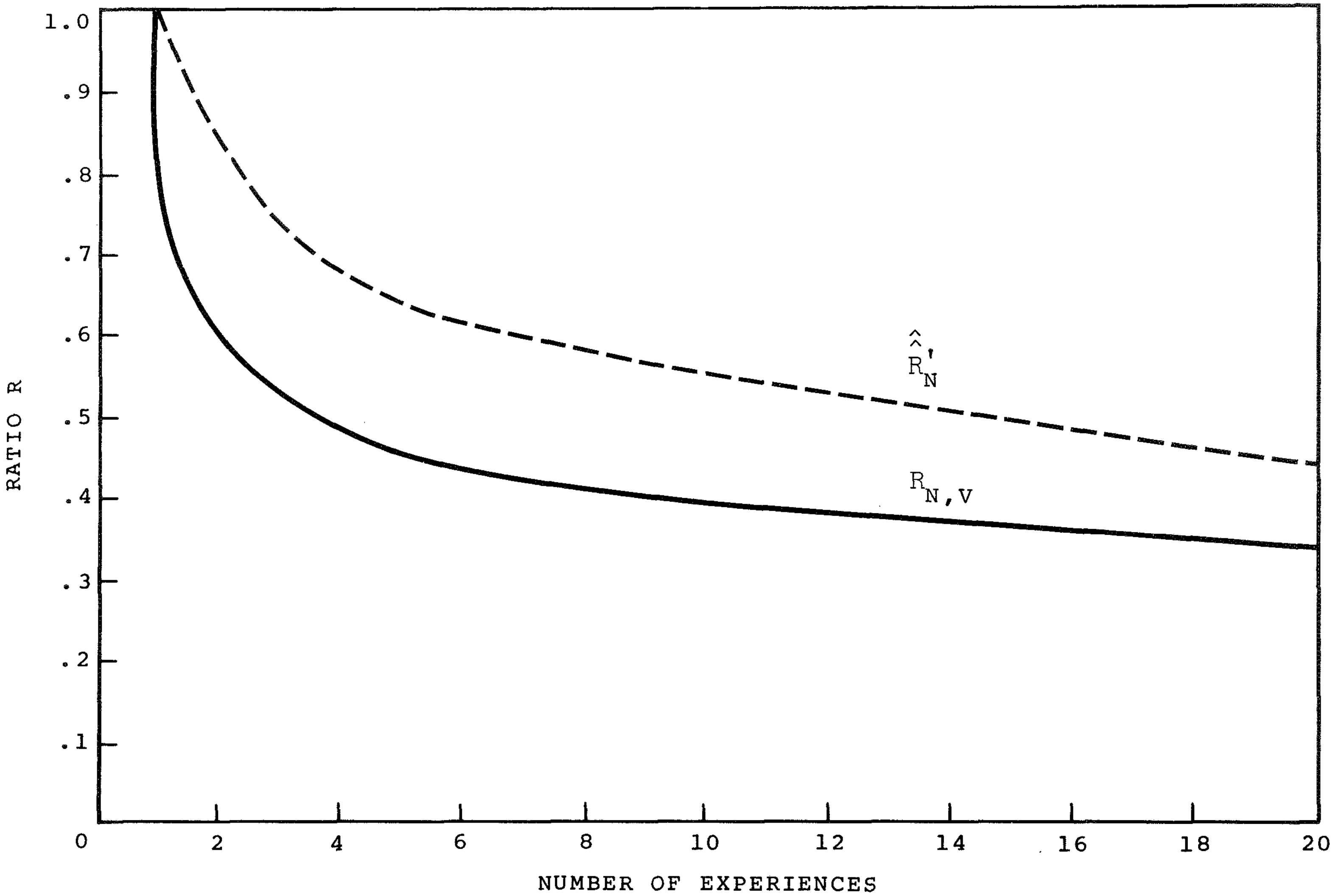

Figure 58. — $\hat{\hat{R}}'_N$ vs $R_{N,V}$ for $Z = 5.0$.

CHAPTER VIII

CONCLUSIONS AND RECOMMENDATIONS

The purpose of this chapter is to summarize the results of this dissertation and to suggest directions for future research.

8.1 General Conclusions

A new method of empirical Bayes estimation has been presented. The versatility of the method has been demonstrated for estimating the parameters of several distributions. The estimator has also been shown to be more efficient in a mean-squared sense than other well-known and widely used empirical Bayes estimators.

The continuously smooth empirical Bayes estimator was developed in Chapter II. The estimator was presented in a general form applicable to multivariate estimation problems. The estimator allows the researcher the use of present as well as previous information, as opposed to classical estimation techniques which must restrict the researcher to the use of only present or current information. The past information is in the form of classical estimates of other parameters in similar but independent experiments. By "similar" we mean that there exists a common prior distribution of the parameter vector, but it remains forever unknown.

The smooth estimator was obtained by representing the prior density function in the Bayes estimator by a continuous approximation formed from a sequence of classical estimates. This approximation was based on three general assumptions. They are as follows:

(i) The prior distribution has a continuous density function.

(ii) The classical estimator is both sufficient and consistent for estimating the parameter vector.

(iii) The distribution of the classical estimator is known.

In practical situations a noncontinuous or discrete prior density would sometimes seem to be an unlikely phenomenon. If, however, the prior density was noncontinuous, perhaps it could be approximated by a continuous density function. The effect such an approximation would have on the smooth estimator is a subject for future research.

In general most classical estimators can be shown to be consistent. In particular, the widely used maximum-likelihood estimators are consistent under general regularity conditions. The assumption of consistency was used to prove that the marginal density function of the classical estimator converges in probability to the prior density function as the sample size tends to infinity. Experimental results revealed that the smooth estimator gave improved results for small sample sizes. Thus it is conjectured that the assumption of consistency represents a property which in practice can be relaxed. If a classical estimator provides "good" results, then a smooth estimator based on it should also give "good" results, regardless of the consistency property. The sufficiency property can also be relaxed. This was demonstrated in Chapters V and VI. Therefore it is conjectured that the second assumption represents an unnecessary restriction in many practical applications.

The third assumption may also be viewed as a general restriction. When the true distribution of the classical estimator is unobtainable, its asymptotic distribution may be known. This distribution can then be used to form a continuously smooth empirical Bayes estimator. In Chapters V and VI this technique was employed, and significant squared-error improvement over the classical maximum-likelihood method was achieved.

In all cases considered in the Monte Carlo studies of the preceding chapters, the continuously smooth empirical Bayes estimators uniformly provided mean-squared improvement over the maximum-likelihood estimators. The smooth estimators were also observed to be robust to the form of the prior distribution. For all types of Pearson prior distributions with varying coefficients of skewness and kurtosis, the ratio R of empirical Bayes mean-squared error to maximum-likelihood mean-squared error was observed to be significantly influenced by the prior distribution only through a quantity Z. Apart from the number of experiences, the only quantities affecting the ratio R are contained in Z. Therefore this quantity was conveniently used to summarize and index the amount of improvement achieved by the smooth estimators over the maximum-likelihood estimators. In particular as the value of Z increased, the ratio R decreased. This phenomenon is easily explained. As the variance of the maximum-likelihood estimator, given the corresponding parameter value, increases relative to variance of the prior distribution, the maximum-likelihood estimates will vary widely. The smooth estimators, however, are capable of "detecting" this variation and can use this information to obtain improved estimates. Conversely, if the conditional variance is small as compared to the prior variance, then the maximum-likelihood estimator would be expected to do quite well. In this case there is a great deal of information within an experiment, and previous experiments contribute very little information about the parameter.

In the preceding chapters whenever point estimates were required for distributions conditional on only one parameter, four smooth estimators were considered. They are: (1) an estimator $\tilde{\Theta}_D$ whose prior density approximation is based on a sequence of classical estimates, (2) an estimator $\tilde{\Theta}'_D$ whose prior density approximation is based on a sequence of smooth estimates obtained from $\tilde{\Theta}_D$, (3) an estimator $\tilde{\Theta}_{D,V}$ whose prior density approximation is based on a sequence of transformed maximum-likelihood estimates having a marginal distribution whose mean and variance are approximately equivalent to those of the prior distribution, and (4) an estimator $\tilde{\Theta}'_{D,V}$ whose prior density approximation is based on a sequence of smooth estimates obtained from $\tilde{\Theta}_{D,V}$. Of these estimators, the smooth estimator $\tilde{\Theta}_{D,V}$ was observed to be the most efficient in a mean-squared sense. This result is expected. The improvement achieved by the smooth estimators was observed to be a function of the mean and variance of the prior distribution as demonstrated by the summary quantity Z; thus any consideration given to accurately estimating these moments by the prior density approximation should result in squared-error improvement.

The transformation given in section 2.5, on which the smooth estimators $\tilde{\theta}_{D,V}$ are based, was only applied to the maximum-likelihood estimators. It is, however, not just restricted to this method of estimation. The transformation can be applied to any classical estimator whose distribution is known.

In Chapter VI smooth estimators were obtained for the parameter vector $\underset{\sim}{\mu} = (\alpha,\beta)$ in the two-parameter Weibull distribution. Two distinct types of smooth estimators were considered. The first type denoted by $\tilde{\alpha}_N$ and $\tilde{\beta}_N$ respectively were constructed in a manner analogous to that given in Chapter II. These estimators were based on an approximation to the bivariate prior density function. No dependence assumptions on α or β were required. The estimators were, however, based on the assumption that the maximum-likelihood estimators are jointly sufficient for α and β . This is not known to be true. The second type of smooth estimators, denoted by $\tilde{\alpha}_M$ and $\tilde{\beta}_M$ respectively and referred to as marginal empirical Bayes estimators, was based on the assumptions that: (1) the maximum-likelihood estimators are independently distributed and are marginally sufficient for α and β respectively, and (2) the parameters α and β are independently distributed. Under these assumptions, the Bayes estimators for α and β become $E(\alpha|\hat{\alpha})$ and $E(\beta|\hat{\beta})$ respectively. In the case of the Weibull distribution, however, neither assumption can generally be made.

In section 6.3 we observed that even when α and β are highly correlated, the estimators $\tilde{\alpha}_M$ and $\tilde{\beta}_M$ gave "better" results than could be obtained using $\tilde{\alpha}_N$ and $\tilde{\beta}_N$. It is conjectured that the compounding effect of being unable to accurately approximate bivariate densities and of having to perform double integration over regions approximated from sample data causes this phenomenon. The error introduced by such approximations appears to be more significant than that caused by the false assumptions of independence and sufficiency.

In the Monte Carlo studies of the preceding chapters, the squared-error improvement achieved by each of the smooth estimators over that of the classical maximum-likelihood estimators was observed to reach its maximum gradient during the initial 10 experiments. For example in Figure 12 for a Z value of 2.0, the smooth estimator

$\tilde{\alpha}_{G,V}$ achieves 34 percent improvement over the maximum-likelihood estimator after the second experience and gains 31 percent more improvement through the tenth experiment. From the tenth to the twentieth experiment, only a 7 percent gradient is noticed. This result tends to indicate that in practice an accumulation of more than 10 sets of data may be unnecessary. The amount of labor required to obtain additional data may not be worth the slight increase in improvement.

In Chapter VII the continuously smooth empirical Bayes estimator was shown to be significantly superior, over 20 experiences, to the estimators proposed by Rutherford and Krutchkoff [28] and the step function estimators evolved from the method of Lemon and Krutchkoff [17]. The estimators of Rutherford and Krutchkoff are ε-asymptotically optimal; however their mean-squared errors appear to be much larger than those of the continuously smooth estimators or the estimators of Lemon and Krutchkoff. It is the author's opinion that although bypassing explicit estimation of the prior distribution may be theoretically desirable as well as convenient, in practical application such procedures are undesirable. As demonstrated in Chapter VII they not only gave poorer results than the explicit estimation procedures, but were limited in application. In particular, the Weibull distribution with unknown shape parameter, considered in Chapter V, could not be placed into any of the families proposed by Rutherford and Krutchkoff. Furthermore, attempts by the author to obtain an empirical Bayes estimator which does not require explicit estimation of the prior distribution were unsuccessful.

The empirical Bayes estimators $\hat{\theta}_D$ and $\hat{\theta}'_D$ suggested by Lemon and Krutchkoff and given by (7.3) and (7.4) respectively were observed to be more efficient than those of Rutherford and Krutchkoff; however they were not as efficient as the continuously smooth estimators. In particular the estimator $\tilde{\theta}_{D,V}$ was observed to be significantly superior, over a run of 20 experiences, to the estimators $\hat{\theta}_D$ and $\hat{\theta}'_D$. It is conjectured that the observed improvements obtained by the smooth estimators result from the continuous prior density approximation being less sensitive to the estimates than is the discrete approximation to the prior distribution.

8.2 Areas for Future Research

Application of the continuously smooth empirical Bayes estimator to multivariate density estimation will be cumbersome if an electronic computer is unavailable.

Even with the aid of a computer the numerical solution of multiple integrals could prove frustrating to the researcher and he may abandon the method for one of lesser precision. Thus a method to avoid this annoying numerical integration would significantly enhance the practical value of the smooth estimation technique. This method would not be generally applicable since it would depend on the kernel of integration. It may be possible, however, to obtain closed-form solutions for the integrals for certain families of distributions. Such solutions may be accomplished by considering various forms for the prior density estimator. The sine function has been chosen in this dissertation, although many other forms are given by Parzen [24] for univariate estimation and by Martz [19] for multivariate estimation. Currently such investigations are being conducted at Texas Tech University, and initial results are encouraging.

Another lucrative area for research lies in Experimental Design. Since empirical Bayes procedures allow the researcher the use of current as well as past information, he should design his experiments with these procedures foremost in mind. Presently, however, such designs are virtually unavailable. The number of experiments and the sample size, which should be used in order to obtain a certain level of performance, are subjects for further research. If such designs could be obtained, then the total number of items subjected to testing procedures could be drastically reduced. For example if the total number of rocket engines subjected to firing tests in an engine development program could be reduced, then a substantial reduction in cost would be realized.

A general field of research would be the application of the smooth estimator to various practical problems. A particularly suitable field of application that is indicated by the research described in this dissertation lies in Reliability and Maintainability Engineering. This suitability is twofold; first, the empirical Bayes approach frequently lends itself to the testing situations encountered in this field, second, the two-parameter Weibull distribution considered in the preceding chapters is a highly versatile and widely used time-to-failure distribution. Presently research of this nature is being conducted at Texas Tech University.

8.3 Conclusion

The purpose of the research described in this dissertation was to develop and exploit a new method of empirical Bayes estimation. This purpose has been accomplished. The continuously smooth estimator has been shown to provide significant improvements over both the classical maximum-likelihood method and other well-known and widely used empirical Bayes methods. In addition, the results of this research constitute a foundation upon which future applications can be based.

LIST OF REFERENCES

1. Bain, L. H., and Antle, C. E. "Estimation of Parameters in the Weibull Distribution," *Technometrics*, Vol. 9, No. 4, p. 621, 1967.

2. Cacoullos, T. "Estimation of a Multivariate Density," *Annals of the Institute of Statistical Mathematics*, Vol. 18, p. 179, 1966.

3. Clemmer, B. A., and Krutchkoff, R. G. "The Use of Empirical Bayes Estimators in a Linear Regression Model," *Biometrika*, Vol. 55, p. 525, 1968.

4. Cohen, C. A., Jr. "Maximum Likelihood Estimation in the Weibull Distribution Based on Complete and on Censored Samples," *Technometrics*, Vol. 7, p. 579, 1965.

5. Cramér, H. *Mathematical Methods of Statistics*. Princeton: Princeton University Press, 1946.

6. Dubey, S. D. "Some Percentile Estimators for Weibull Parameters," *Technometrics*, Vol. 9, p. 119, 1967.

LIST OF REFERENCES (Continued)

7. Dubey, S. D. "On Some Statistical Inferences for Weibull Laws," *Naval Research Logistics Quarterly*, Vol. 13, p. 227, 1965.

8. Fisher, R. A., and Tippett, L. H. C. "Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample," *Proc. Cambridge Phil. Soc.*, Vol. 24, p. 180, 1928.

9. Govindarajulu, Z., and Joshi, M. "Best Linear Unbiased Estimation of Location and Scale Parameters of Weibull Distribution Using Ordered Observations," Report No. 102, Statistical Laboratory, Case Institute of Technology, Cleveland, Ohio, 1962.

10. Gumbel, E. J. *Statistics of Extremes*. New York: Columbia University Press, 1958.

11. Harris, C. M., and Singpurwalla, N. D. "Life Distributions Derived from Stochastic Hazard Functions." *IEEE Transactions on Reliability*, Vol. 17, p. 70, 1968.

LIST OF REFERENCES (Continued)

12. Harter, H. L., and Moore, A. H. "Maximum-Likelihood Estimation of Parameters of Gamma and Weibull Populations from Complete and Censored Samples," *Technometrics*, Vol. 7, p. 639, 1966.

13. Hildebrand, F. B. *Introduction to Numerical Analysis*. New York: McGraw-Hill Book Co., 1956.

14. Hoel, P. G. *Introduction to Mathematical Statistics*. New York: John Wiley and Sons, Inc., 1966.

15. Kao, J. H. K. "Computer Methods for Estimating Weibull Parameters in Reliability Studies," Technical Report No. 34, Industrial Engineering Dept., Cornell University, New York, 1957.

16. Kendall, M. G., and Stuart, A. *The Advanced Theory of Statistics*, Vol. 2, New York: Hafner Publishing Company, 1961.

17. Lemon, G. H., and Krutchkoff, R. G. "An Empirical Bayes Smoothing Technique," *Biometrika*, Vol. 56, p. 361, 1969.

LIST OF REFERENCES (Continued)

18. Lowan, A. N., Davids, N., and Levenson, A. "Tables of Zeros of the Legendre Polynomials of Order 1-16 and the Weight Coefficients for Gauss Mechanical Quadrature Formula," *Bulletin of the American Mathematical Society*, Vol. 48, p. 739, 1942.

19. Martz, H. F., Jr. "Empirical Bayes Estimation in Multiple Linear Regression," Unpublished Ph.D. Dissertation, Virginia Polytechnic Institute, 1967.

20. Martz, H. F., Jr., and Krutchkoff, R. G. "Empirical Bayes Estimators in a Multiple Linear Regression Model," *Biometrika*, Vol. 56, p. 367, 1969.

21. Menon, M. V. "Estimation of the Shape and Scale Parameters of the Weibull Distribution," *Technometrics*, Vol. 5, p. 175, 1963.

22. Miller, I., and Freund, J. E. *Probability and Statistics for Engineers*. New Jersey: Prentice-Hall Inc., 1965.

LIST OF REFERENCES (Continued)

23. Mood, A. M., and Graybill, F. A. *Introduction to the Theory of Statistics*. New York: McGraw-Hill Book Co., 1963.

24. Parzen, E. "On Estimation of a Probability Density Function and Mode," *Ann. Math. Statist.*, Vol. 33, p. 1065, 1962.

25. Rao, C. R. *Linear Statistical Inference and Its Applications*. New York: John Wiley and Sons, Inc., 1965.

26. Robbins, H. E. "An Empirical Bayes Approach to Statistics," *Proc. Third Berkeley Symp.*, Vol. 1, p. 157, 1955.

27. Rutherford, J. R. "Some Parametric Empirical Bayes Techniques," Unpublished Ph.D. Dissertation, Virginia Polytechnic Institute, 1965.

28. Rutherford, J. R., and Krutchkoff, R. G. "Some Empirical Bayes Techniques in Point Estimation," *Biometrika*, Vol. 56, p. 221, 1969.

LIST OF REFERENCES (Continued)

29. Rutherford, J. R., and Krutchkoff, R. G. "ε Asymptotic Optimality of Empirical Bayes Estimators," *Biometrika*, Vol. 56, p. 220, 1964.

30. Soland, R. M. "Bayesian Analysis of the Weibull Process with Unknown Scale Parameter and its Application to Acceptance Sampling," *IEEE Transactions on Reliability*, Vol. 17, p. 84, 1968.

31. Thoman, D. R., Bain, L. J., and Antle, C. E. "Inferences on the Parameters of the Weibull Distribution," *Technometrics*, Vol. 11, p. 445, 1969.

32. Wald, A. "Note on the Consistency of the Maximum Likelihood Estimate," *Ann. Math Statist.*, Vol. 20, p. 595, 1949.

33. White, J. S. "Least-Squares Unbiased Censored Linear Estimation for the Log Weibull (Extreme-Value) Distribution," *J. Industrial Math. Soc.*, Vol. 14, p. 21, 1964.

LIST OF REFERENCES (Continued)

34. Weibull, W. "A Statistical Theory of the Strength of Material," *Ing. Vetenskaps Akad. Handl.*, Vol. 153, 1939.

35. Weibull, W. "The Phenomenon of Rupture in Solids," *Ing. Vetenskaps Akad. Handl.*, Vol. 151, 1939.

36. Weibull, W. "A Statistical Distribution of Wide Applicability," *J. Appl. Mech.*, Vol. 18, p. 293, 1951.

37. Weibull, W. "Statistical Design of Fatigue Experiments," *J. Appl. Mech.*, Vol. 19, p. 109, 1952.

APPENDIX

A. Gauss Quadrature

B. Newton's Method

C. Documentation of Computer Programs Used in the Monte Carlo Simulation for the Poisson Distribution

D. Documentation of Computer Programs Used in the Monte Carlo Simulation for the Weibull Distribution with Unknown Scale Parameter α

E. Documentation of Computer Programs Used in the Monte Carlo Simulation for the Weibull Distribution with Unknown Shape Parameter β

F. Documentation of Computer Programs Used in the Monte Carlo Simulation for the Weibull Distribution

APPENDIX A: GAUSS QUADRATURE

An accurate formula for finding the value of the definite integral

$$I = \int_a^b f(x)\, dx \qquad (A.1)$$

where f(x) is a known function, but whose integral is either not easily evaluated or cannot be conveniently expressed in closed form, was derived by Gauss and is based on Legendre polynomials. The procedure is to obtain the subdivision of the interval (a,b), the value of the function at these points, and the coefficients to multiply the functional values to yield the value of the definite integral.

First we transform the interval $x = (a,b)$ into the interval $t = (-1,1)$ by letting

$$x = \frac{1}{2}(b - a)t + \frac{1}{2}(a + b) \quad .$$

The new form of f(x) is

$$f(x) = f\left[\frac{1}{2}(b - a)t + \frac{1}{2}(a + b)\right] = \Phi(t)$$

and

$$dx = \frac{1}{2}(b - a)\, dt$$

so that

$$\int_a^b f(x)\, dx = \frac{b - a}{2} \int_{-1}^{1} \Phi(t)\, dt \quad .$$

Using the Gauss mechanical quadrature formula

$$\int_{-1}^{1} \Phi(t)\, dt = \sum_{k=1}^{n} \left[A_k^{(n)} \Phi\left(t_k^{(n)}\right) \right] , \qquad (n = 2,3,\cdots)$$

where n is the number of points of subdivision of the interval $(-1,1)$, $A_k^{(n)}$ the weighting coefficients, and $t_k^{(n)}$ the zeros of the Legendre polynomials of degree n , we have

$$I_n = \frac{b - a}{2} \sum_{k=1}^{n} A_k^{(n)} f\left[(b - a) \frac{t_k^{(n)}}{2} + \frac{b + a}{2} \right]$$

$$(n = 2,3,\cdots) \quad .$$

The $A_k^{(n)}$ and $t_k^{(n)}$ are symmetric with respect to $t = 0$, that is,

$$A_k^{(n)} = A_{n-k+1}^{(n)} , \qquad t_k^{(n)} = - t_{n-k+1}^{(n)} \quad .$$

A table of the zeros $t_k^{(n)}$ of the Legendre polynomial of order 1-16 and the weight coefficients $A_k^{(n)}$ for the Gauss mechanical quadrature are given by Lowan et al. [18]. In Table A1 we have reproduced these values for the special cases when $n = 3$ and $n = 11$.

In particular, in the subroutine GAUSS, which was used to solve the integrals of the smooth estimator, n was taken to be 11. Also a "built in" accuracy check is made. The value of the integral given by (A.1) is first calculated; then the integral

$$I' = \int_a^{\frac{b-a}{2}} f(x)\,dx + \int_{\frac{b-a}{2}}^{b} f(x)\,dx$$

is computed. If $\delta < \varepsilon$, where

$$\delta = \begin{cases} \left|\dfrac{I' - I}{I'}\right| & \text{if } |I'| > \varepsilon \\ |I' - I| & \text{if } |I'| \leq \varepsilon \end{cases}$$

and ε is the desired tolerance given by input, then the value I' is returned as the value of the integral. If $\delta \geq \varepsilon$ then the range (a,b) is partitioned into four subintervals and the sum I'' of the integrals over each interval is computed and δ is formed with I''

TABLE A1

GAUSS'S QUADRATURE COEFFICIENTS

k	t_k	A_k
	$n = 3$	
1	0.77459 66692	0.55555 55556
2	0.00000 00000	0.88888 88889
	$n = 11$	
1	0.97822 86581	0.05566 85671
2	0.88706 25998	0.12558 03695
3	0.73015 20056	0.18629 02109
4	0.51909 61291	0.23319 37646
5	0.26954 31560	0.26280 45445
6	0.00000 00000	0.27292 50868

and I'. If $\delta < \varepsilon$ then I'' is returned as the value of the integral. If $\delta \geq \varepsilon$ then the region (a,b) is again subdivided. This procedure is repeated IT times, a value given by input. If convergence is not reached after IT subdivisions, an error message is printed and the program terminated.

APPENDIX B: NEWTON'S METHOD

The well-known iterative procedure of Newton can be easily applied to the solution of the nonlinear maximum-likelihood estimating equations of Chapters V and VI. This procedure can be obtained by truncating the Taylor series expansion after two terms and has the form

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} \qquad (i = 0,1,\cdots) \quad . \qquad (B.1)$$

Convergence to the root is quadratic if the multiplicity of the root to be determined is equal to one and if f(x) is a twice-differentiable function.

The subroutine RTNI based on (B.1) and given in the System/360 Scientific Subroutine Package was used to solve for the maximum-likelihood estimators. In this subroutine the iterative procedure is terminated if the following two conditions are satisfied:

$$\delta \leq \varepsilon \quad \text{and} \quad |f(x_i + 1)| \leq 100\varepsilon$$

with

$$\delta = \begin{cases} \left| \dfrac{x_{i+1} - x_i}{x_{i+1}} \right| & \text{if} \quad |x_{i+1}| > 1 \\ \\ |x_{i+1} - x_i| & \text{if} \quad |x_{i+1}| \leq 1 \end{cases}$$

and tolerance ε given by input. If the procedure does not converge within a specified number of iteration steps, an error message is given. For further details on the method as well as reasons for divergence, see Hildebrand [13].