Maximum Likelihood Estimation of the Parameters of the Beta Distribution from Smallest Order Statistics

R. GNANADESIKAN, R. S. PINKHAM* AND LAURA P. HUGHES

Bell Telephone Laboratories, Incorporated

Numerical methods, useful with high-speed computers, are described for obtaining the maximum likelihood estimates of the two (shape) parameters of a beta distribution using the smallest M order statistics, $0 < u_1 \leq u_2 \leq \cdots \leq u_M$, in a random sample of size $K(\geq M)$. The maximum likelihood estimates are functions only of the ratio, $R = M/K$, the Mth ordered observation, u_M, and the two statistics, $G_1 = [\Pi_{i=1}^{M} u_i]^{1/M}$ and $G_2 = [\Pi_{i=1}^{M} (1 - u_i)]^{1/M}$. For the case of the complete sample (i.e., $R = 1$), however, the estimates are functions only of G_1 and G_2, and hence, for this case, explicit tables of the estimates are provided. When $R < 1$, the methods described depend crucially for their usefulness on the availability of a high-speed computer.

Some examples are given of the use of the procedures described for fitting beta distributions to sets of data. In one example, the fit is studied by using beta probability plots.

1. INTRODUCTION

The family of beta distributions is related to many of the common statistical distributions, including the t, F, binomial and negative binomial distributions. Also, the beta distribution has been used in certain Bayesian applications as a prior distribution for the binomial parameter, p. [See, for example, Anscombe (1961).] Chaddha has, in some unpublished reports, suggested the use of a special case of the beta distribution as a model in Queueing Theory and in reliability applications. The beta distribution may also serve as an appropriate approximation to fit the distribution of the probability integral transformation, when estimates of parameters are used in the transformation so that the transformed variable may no longer have the uniform distribution. [See David and Johnson (1948).]

The present paper is concerned with the maximum likelihood estimation of the two parameters of an underlying beta distribution using the smallest observations in a random sample. The case of maximum likelihood estimation from the complete sample is included as a special case.

The formulation of the problem in terms of smallest order statistics appears to be natural for reliability applications where the data often arrive in an ordered fashion starting with the smallest observation. Also, in many applications

Received June 1966; revised February 1967.

* Professor R. S. Pinkham is now at Stevens Institute of Technology.

of probability plotting, such as in the analysis of variance, when the data to be plotted is also to provide the basis for estimating the necessary parameters, a reasonable and natural formulation of the estimation problem appears to be one in terms of smallest order statistics. [cf., for example, Wilk and Gnanadesikan (1961, 1964) and Wilk et al. (1963).] In other applications, one may wish to base the estimation on the "middle" or "larger" observations in the sample. [cf. Wilk et al. (1966).]

The results of the present paper are germane to an interest in fitting beta distributions. Section 2 of the paper gives a statement of the problem with necessary notation. In Section 3 the likelihood equations are derived and in Section 4 numerical methods employed in solving these are described. Some examples of application of the estimation procedure are given in Section 5. Section 6 consists of summary remarks and general discussion. An appendix contains the numerical approximations used in solving the likelihood equations. A table of the maximum likelihood estimates for the complete sample case is included in Section 3.

2. Notation and Statement of the Problem

Consider the ordered observations, $0 < u_1 \le u_2 \le \cdots \le u_K < 1$, resulting from a random sample of size K from a beta distribution with density

$$f(u; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} u^{\alpha-1}(1-u)^{\beta-1}, \qquad 0 < u < 1, \qquad \alpha > 0, \qquad \beta > 0. \quad (1)$$

It is desired to estimate α and β simultaneously, utilizing the $M(\le K)$ smallest observations, $u_1, u_2, \cdots, u_M$. The method of estimation used in the sequel is that of maximizing the likelihood based on this formulation in terms of order statistics.

The estimation problem for the so-called beta type II distribution with density

$$g(v; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} \frac{v^{\alpha-1}}{(1+v)^{\alpha+\beta}}, \qquad 0 \le v < \infty, \qquad \alpha > 0, \qquad \beta > 0,$$

can be transformed to the above framework by setting $u = v/(1+v)$.

3. Maximum Likelihood Estimation

The likelihood of α and β given the M smallest observations, $u_1, u_2, \cdots, u_M$ is

$$\mathcal{L}(\alpha, \beta) \propto \left[\frac{1}{B(\alpha, \beta)}\right]^K \left[\prod_{i=1}^{M} u_i^{\alpha-1}(1-u_i)^{\beta-1}\right]\left[\int_{u_M}^{1} t^{\alpha-1}(1-t)^{\beta-1}\, dt\right]^{K-M}. \quad (2)$$

The likelihood equations, $\partial \log \mathcal{L}/\partial\alpha = 0$ and $\partial \log \mathcal{L}/\partial\beta = 0$, reduce to

$$R \ln G_1 = \Psi(\alpha) - \Psi(\alpha+\beta) - (1-R)\frac{I_1(u_M; \alpha, \beta)}{I(u_M; \alpha, \beta)} \quad (3)$$

and

$$R \ln G_2 = \Psi(\beta) - \Psi(\alpha + \beta) - (1 - R) \frac{I_2(u_M ; \alpha, \beta)}{I(u_M ; \alpha, \beta)}, \tag{4}$$

where

$$R = M/K, \qquad G_1 = \left(\prod_{i=1}^{M} u_i\right)^{1/M}, \qquad G_2 = \left(\prod_{i=1}^{M} (1 - u_i)\right)^{1/M},$$

$$\Psi(x) = \frac{d}{dx} \ln \Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)},$$

$$I(x; \alpha, \beta) = \int_x^1 t^{\alpha-1}(1 - t)^{\beta-1}\, dt, \qquad (0 \leq x \leq 1),$$

$$I_1(x; \alpha, \beta) = \frac{\partial}{\partial\alpha} I(x; \alpha, \beta) = \int_x^1 t^{\alpha-1}(1 - t)^{\beta-1} \ln t\, dt, \qquad (0 \leq x \leq 1),$$

and

$$I_2(x; \alpha, \beta) = \frac{\partial}{\partial\beta} I(x; \alpha, \beta) = \int_x^1 t^{\alpha-1}(1 - t)^{\beta-1} \ln (1 - t)\, dt, \qquad (0 \leq x \leq 1).$$

Clearly the maximum likelihood estimates, α and β, depend on u_M, G_1, G_2 and R. In general, therefore, tabulation of the roots of equations (3) and (4) would involve four-way tables which would be too unwieldy for practical purposes. However, in the special case when $R = 1$, (i.e., the complete sample case), the likelihood equations simplify to

$$\ln G_1 = \Psi(\alpha) - \Psi(\alpha + \beta), \tag{5}$$

$$\ln G_2 = \Psi(\beta) - \Psi(\alpha + \beta). \tag{6}$$

A tabulation of the roots of these equations, in terms of (observed values) G_1 and G_2, is provided in Table 1. While the maximum likelihood estimates for specific values of G_1 and G_2, and some indication of the general pattern of their values, may be gleaned from Table 1, yet the grid shown for G_1 and G_2 is so coarse that linear interpolation in the table would yield accuracies of at most only two significant digits and not even single-significant-digit accuracy when G_1 and G_2 have extremely disparate values. More detailed tables, using a grid for G_1 and G_2 which is fine enough for linear interpolation to be adequate for three-significant-digit accuracy, are available but have not been included here. The problem of tabulating values of some simple single-valued functions of the estimates, which are more nearly linear in G_1 and G_2, instead of the estimates themselves is under continuing study.

The singularities in the equations when either G_1 or G_2 is zero, led to considering 0.01 as the smallest value of G_1 and G_2 for tabulation purposes. Further, $G_1 + G_2 \leq 1$, with the equality holding if and only if the observations are all equal which is an unlikely and uninteresting occurrence in practice. Hence, the largest values of G_1 (or G_2) for a specified value of $G_2(G_1)$ was chosen such that $G_1 + G_2 = 0.99$. The condition, $G_1 + G_2 \leq 1$, and the symmetry implied

TABLE 1

Complete sample maximum likelihood estimates of parameters of the beta distribution. Range of G_1 is .01, .1(.1).9. Range of G_2 is .01, .1(.1).9. $[G_1 + G_2 \leq 1.0]$ In each cell the first entry is $\hat{\alpha}$ and the second entry is $\hat{\beta}$.

G_1 / G_2	.01	.1	.2	.3	.4	.5	.6	.7	.8	.9	.98
.01	.112 .112	.192 .135	.254 .147	.318 .157	.395 .168	.495 .179	.639 .192	.877 .210	1.376 .237	3.162 .299	42.128 .850
.1		.245 .245	.337 .278	.441 .310	.576 .345	.770 .389	1.093 .451	1.756 .560	3.864 .846	* *	
.2			.395 .395	.537 .456	.735 .531	1.057 .640	1.701 .834	3.669 1.367	* *		
.3				.647 .647	.947 .804	1.532 1.086	3.280 1.869	* *			
.4					1.320 1.320	2.832 2.358	* *				
.5	BY SYMMETRY**					* *					
.6					* *						
.7				* *			NOT POSSIBLE $(G_1 + G_2 > 1)$				
.8			* *								
.9		* *									
.98											

* $G_1 + G_2 = 1$

** For example, the estimates of α and β when $G_2 = 0.2$ and $G_1 = 0.1$ are respectively the estimates of β and α when $G_1 = 0.2$ and $G_2 = 0.1$, namely 0.278 and 0.337.

by equations (5) and (6) with respect to G_1 and G_2 lead to Table 1 having fewer entries than might be anticipated.

4. Method for Solving Likelihood Equations

The expressions on the right-hand sides of the likelihood equations (3) and (4) may be denoted $F_1(\alpha, \beta)$ and $F_2(\alpha, \beta)$ respectively, and the likelihood equations rewritten as

$$R \ln G_1 = F_1(\alpha, \beta), \tag{7}$$

$$R \ln G_2 = F_2(\alpha, \beta). \tag{8}$$

F_1 and F_2 involve R and u_M in addition to being functions of α and β. Given the sample of observations to be used in the estimation of α and β, the quantities

R, u_M, G_1 and G_2 are known, and the aim is to find the values of α and β that satisfy equations (7) and (8) simultaneously. Knowing R and u_M, the functions F_1 and F_2 can be evaluated at any specified point, (α_0, β_0), provided that Ψ, I, I_1, and I_2, defined in Section 3, can be evaluated at (α_0, β_0). Approximations useful for computing these functions are given in the Appendix and a detailed discussion of them and the attendant errors may be found in Gnanadesikan et al. (1966). Using these approximations and the given values of R and u_M, one can compute $F_1(\alpha_0, \beta_0)$ and $F_2(\alpha_0, \beta_0)$ at a trial value, (α_0, β_0), for the root, $(\hat{\alpha}, \hat{\beta})$, of equations (7) and (8).

Iterative methods may be employed for the numerical solution of the likelihood equations, (7) and (8). Newton's method, involving the linearization of F_1 and F_2 in the neighborhood of the root, leads to an iterative scheme defined by

$$\left.\begin{aligned} R \ln G_1 &= F_1(\alpha_n, \beta_n) + (\alpha_{n+1} - \alpha_n)\left(\frac{\partial F_1}{\partial \alpha}\right)_{\alpha_n, \beta_n} + (\beta_{n+1} - \beta_n)\left(\frac{\partial F_1}{\partial \beta}\right)_{\alpha_n, \beta_n} \\ R \ln G_2 &= F_2(\alpha_n, \beta_n) + (\alpha_{n+1} - \alpha_n)\left(\frac{\partial F_2}{\partial \alpha}\right)_{\alpha_n, \beta_n} + (\beta_{n+1} - \beta_n)\left(\frac{\partial F_2}{\partial \beta}\right)_{\alpha_n, \beta_n} \end{aligned}\right\} \quad n = 0, 1, 2, \cdots \tag{9}$$

Instead of using the explicit, computable, analytical approximations, provided by Gnanadesikan et al. (1966), for the partial derivatives of F_1 and F_2 with respect to α and β, the present approach employed the oft used device of approximating derivatives by divided differences. Thus, the iterative scheme defined by equation (9) was used with the modification that the partial derivatives are replaced by

$$\left.\begin{aligned} \frac{\Delta F_i}{\Delta \alpha} &= \frac{F_i(\alpha^*, \beta_n) - F_i(\alpha_n, \beta_n)}{\alpha^* - \alpha_n}, \\ \text{and} \qquad \frac{\Delta F_i}{\Delta \beta} &= \frac{F_i(\alpha_n, \beta^*) - F_i(\alpha_n, \beta_n)}{\beta^* - \beta_n}, \end{aligned}\right\} \quad i = 1, 2. \tag{10}$$

The mesh sizes, $\alpha^* - \alpha_n$ and $\beta^* - \beta_n$, which are essential ingredients of the approximation of the derivatives by the divided differences, were chosen to be both equal to 10^{-3} based on preliminary empirical investigations.

The starting values, (α_0, β_0), used are crucial for the efficient convergence of the iterative scheme. When $R = 1$, the moment estimates provide an adequate set of starting values defined by

$$\alpha_0 = \frac{m_1(m_1 - m_2)}{m_2 - m_1^2}, \qquad \beta_0 = \frac{1 - m_1}{m_1} \alpha_0, \tag{11}$$

where $m_1 = (1/K) \sum_{i=1}^{K} u_i$, and $m_2 = (1/K) \sum_{i=1}^{K} u_i^2$. When $R < 1$, however, the moment estimates from the observations on hand, $u_1, u_2, \cdots, u_M$, appear to be inappropriate and inadequate for obtaining starting values that lead to rapid convergence. For this case, therefore, alternate starting values were devised. The basic idea consists of equating certain of the order statistics to quantiles of an approximating distribution suggested by Wise (1950). Specifically, Wise (1950, equation 5.2) suggests, as a first order approximation for the quantile, x_p, of a beta distribution

$$\ln x_p = -\chi^2_{2\beta}(p)/2N, \tag{12}$$

where $\chi^2_{2\beta}(p)$ denotes the quantile corresponding to p, $(0 < p < 1)$, of a χ^2 distribution with 2β degrees of freedom, and $N = \alpha + (\beta - 1)/2$. Thus, corresponding to the order statistics, u_1 and u_M, one obtains from equation (12), the relationships

$$\ln u_1 \simeq -\frac{\chi^2_{2\beta}(1/K+1)}{2N}, \qquad \ln u_M \simeq -\frac{\chi^2_{2\beta}(M/K+1)}{2N}. \tag{13}$$

As a further simplification, the well-known standard normal approximation to the distribution of $(\sqrt{2\chi^2_\nu} - \sqrt{2\nu - 1})$ was used with the relationships provided by equation (13), and the following starting values, α_0 and β_0, were obtained:

$$\beta_0 = \tfrac{1}{4}(\beta^{*2} + 1), \qquad \alpha_0 = N_0 - \tfrac{1}{2}(\beta_0 - 1), \tag{14}$$

where

$$\beta^* = \left[\left\{\frac{\ln u_1}{\ln u_M}\right\}^{\frac{1}{2}} z_{(M/K+1)} - z_{(1/K+1)}\right] \Big/ \left[1 - \left\{\frac{\ln u_1}{\ln u_M}\right\}^{\frac{1}{2}}\right],$$

and

$$N_0 = -\frac{\chi^2_{2\beta_0}(1/K+1)}{2 \ln u_1},$$

and $z_{(p)}$ denotes the standard normal quantile corresponding to the proportion p. Any of the available methods for computing the percentage points of a chi-square distribution (see for example, Wilk et al. (1962a)) might be used to obtain $\chi^2_{2\beta_0}(1/K + 1)$ and thence N_0. For $R < 1$, the starting values provided by equation (14) were used in the iterative scheme defined by equation (9) and the iterations repeated until $|R \ln G_1 - F_1(\alpha_n, \beta_n)| \leq 10^{-4}$ and $|R \ln G_2 - F_2(\alpha_n, \beta_n)| \leq 10^{-4}$. The values, α_n, β_n, at the first stage of iteration when these criteria of convergence are met are then the desired maximum likelihood estimates, $\hat{\alpha}$ and $\hat{\beta}$.

For the desired degree of convergence, the authors found, in some cases, that the number of iterations required was fairly large. In these cases, the pattern of convergence was one wherein within a few iterations (i.e., for small n), either $|R \ln G_1 - F_1(\alpha_n, \beta_n)| \leq 10^{-4}$ or $|R \ln G_1 - F_2(\alpha_n, \beta_n)| \leq 10^{-4}$, but not both criteria were satisfied simultaneously for small n. To cut the number of iterations (and the saving was found to be considerable in many cases), the authors used a scheme for adjusting the values of α_i, β_i at certain stages of the iteration. For instance, if F_1 is adequately close to $R \ln G_1$ and has remained thus for a specified number of iterations (the value used was 5) including the ith iteration, while F_2 is not sufficiently close to $R \ln G_2$, then α_{i+1} and β_{i+1} were adjusted by the value of the ratio of $R \ln G_2$ to the computed value of $F_2(\alpha_i, \beta_i)$. That is, $\alpha_{i+1} = \alpha_i R \ln G_2/F_2(\alpha_i, \beta_i)$ and $\beta_{i+1} = \beta_i R \ln G_2/F_2(\alpha_i, \beta_i)$. Similarly, if F_2 is adequately close to $R \ln G_2$ and has remained thus for five iterations including the ith one, while F_1 is not sufficiently close to $R \ln G_1$, then $\alpha_{i+1} = \alpha_i R \ln G_1/F_1(\alpha_i, \beta_i)$ and $\beta_{i+1} = \beta_i R \ln G_1/F_1(\alpha_i, \beta_i)$.

5. Examples of Application

The methods of Sections 3 and 4 are next illustrated by application to two sets of data.

The first set consists of a computer-generated random sample of size $K = 20$ from a beta distribution with parameters $\alpha = 1.5$ and $\beta = 11.0$. Table 2 gives

Table 2*

Maximum Likelihood Estimation of Parameters of the Beta Distribution from Order Statistics Generated Data where Alpha is 1.5 and Beta is 11.0

Sample Size = 20		No Probability Plots		
Observation Number	Ordered Observation	Fraction of Sample	Alpha Estimate	Beta Estimate
1	0.15396729E−01			
2	0.32086748E−01	0.100	1.443	6.472
3	0.40187541E−01			
4	0.45033980E−01	0.200	1.472	8.145
5	0.47815502E−01			
6	0.52427629E−01	0.300	1.483	9.602
7	0.79288867E−01			
8	0.86755657E−01	0.400	1.564	9.760
9	0.89401839E−01			
10	0.90071268E−01	0.500	1.672	11.765
11	0.10152799E−00			
12	0.10534459E−00	0.600	1.774	13.175
13	0.10610413E−00			
14	0.11928693E−00	0.700	1.902	14.995
15	0.18714180E−00			
16	0.19774591E−00	0.800	1.732	12.202
17	0.20310399E−00			
18	0.23729337E−00	0.900	1.800	12.920
19	0.30387626E−00			
20	0.31532391E−00	1.000	1.793	12.781

* *The titles and format of this table are reproduced from computer output.*

the ordered sample and contains the maximum likelihood estimates of α and β for $R = 0.1(0.1)1$. When $R = 0.1$, in this case, only two ordered observations are being used to estimate the two parameters. The indications from this example, which are typical, are that the maximum likelihood estimates are statistically reasonably behaved with biases being negative for small R, positive for large R and with the estimates of the two parameters tending to be positively correlated. In this example, the number of iterations involved in obtaining the solutions of the likelihood equations ranged from 7 to 19, the former being for $R = 1$ and the latter for $R = 0.4$.

The second set of data arose in a problem of talker identification (Becker et al. (1965)) where there was an interest in selecting features whose variations across speakers relative to variations within speakers are large. The data consisted of ratios of "between-speakers" mean squares to "within-speakers" mean squares for a specific set of 20 statistics that were of interest. These ratios of mean squares were transformed to the "equivalent" ratios, $b/(b + w)$, with range 0 to 1 where b and w denote, respectively, the sums of squares between and within speakers. The resulting data, ordered, is given in Table 3 which also contains the estimates obtained corresponding to four values of R.

The estimates when R is 0.25, 0.5 and 0.75 are quite similar while being quite different from those when the complete sample ($R = 1$) is used. In the present application the larger observations in the sample might indeed depart markedly from the statistical configuration of the rest of the sample and the estimates might, for some purposes, be based more reasonably on $R < 1$ rather than on $R = 1$.

In order to study further the implications of these different sets of estimates and to uncover possible peculiarities in the sample (e.g., some observations being overly large), the sets of estimates when $R = 0.5$ and when $R = 1$ were used and probability plots of the data obtained. These are plots of the ordered data against appropriate quantiles of the fitted beta distribution and their use and interpretation is quite similar to those of other probability plots. (See, for example, Wilk et al. (1962a), Wilk and Gnanadesikan (1964).)

TABLE 3*

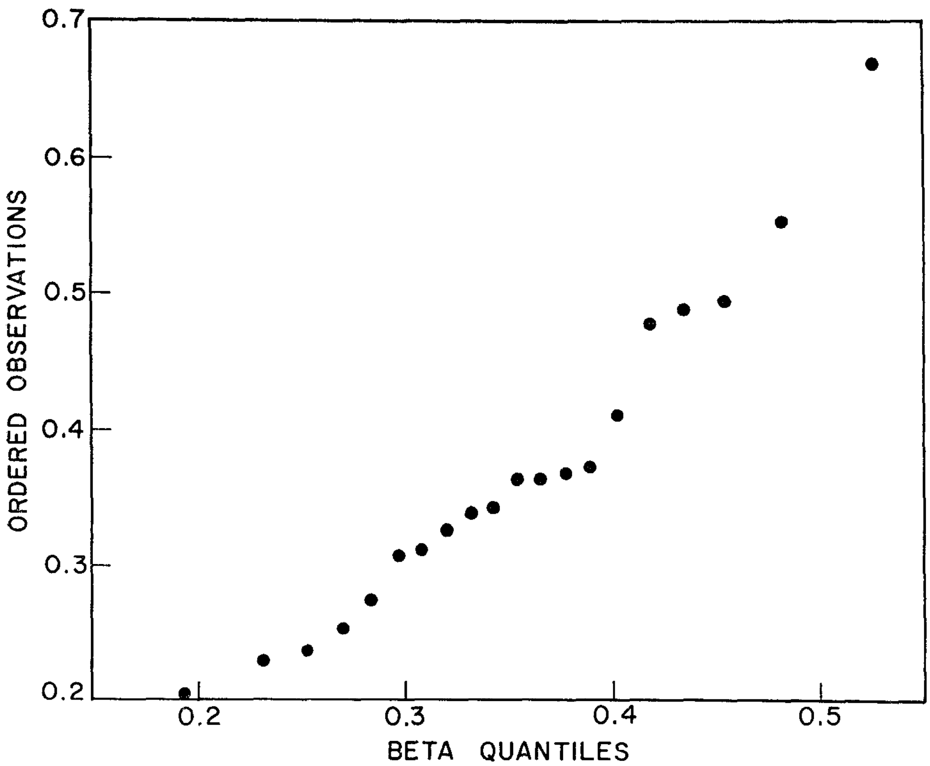
Maximum Likelihood Estimation of Parameters of the Beta Distribution from Order Statistics Data from Talker Identification Problem

Sample Size = 20		Probability Plots Included		
Observation Number	Ordered Observation	Fraction of Sample	Alpha Estimate	Beta Estimate
1	0.20272330E−00			
2	0.23144300E−00			
3	0.23827390E−00			
4	0.25358830E−00			
5	0.27574400E−00	0.250	9.565	18.076
6	0.30974110E−00			
7	0.31187140E−00			
8	0.32694060E−00			
9	0.33953980E−00			
10	0.34379880E−00	0.500	10.479	19.378
11	0.36592120E−00			
12	0.36616030E−00			
13	0.36928880E−00			
14	0.37460130E−00			
15	0.41196090E−00	0.750	10.377	19.014
16	0.47984940E−00			
17	0.48932910E−00			
18	0.49532190E−00			
19	0.55465619E+00			
20	0.66715830E+00	1.000	6.543	11.052

* *Format reproduced from computer output.*

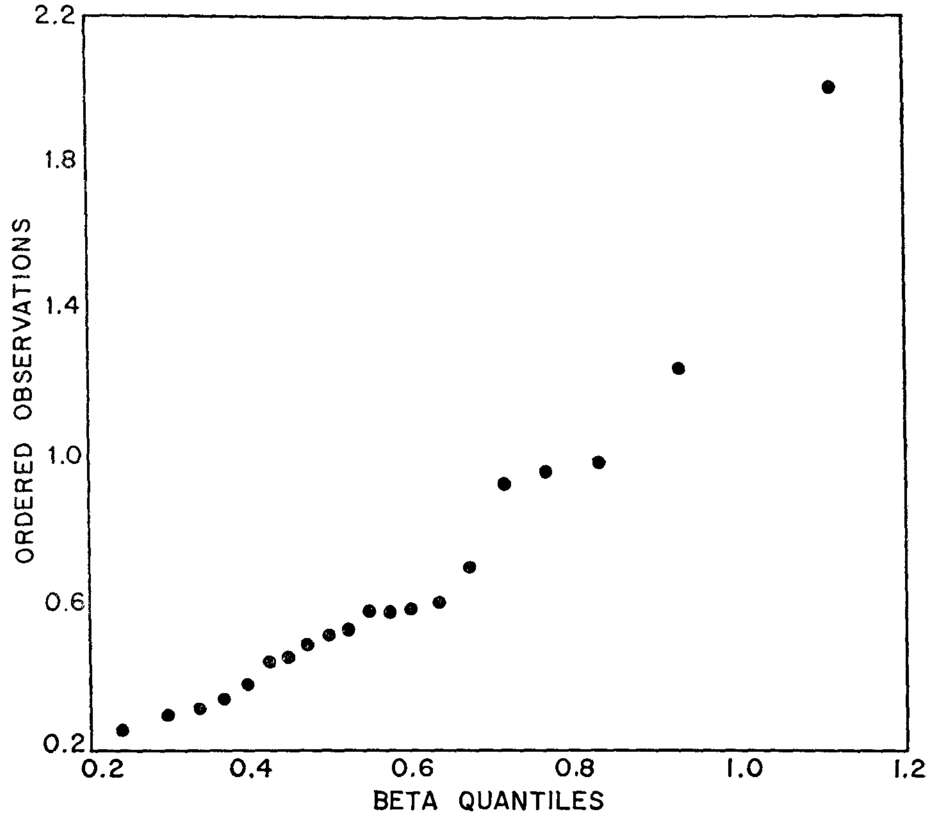
For each set of estimates, $\hat{\alpha}$ and $\hat{\beta}$, two types of probability plots were obtained. The first corresponds to plotting the ordered quantities with range 0 to 1 against quantiles of the fitted beta type I distribution while the second is a plot of the ordered original ratios of mean squares against quantiles of the fitted beta type II distribution. Figure A_1 shows the beta type I plot based on the estimates obtained when $R = 0.5$, and Figure A_2 shows the corresponding beta type II plot. In spite of the mildly ragged nature of the plots, the "largest point" (i.e., the one in the top right-hand corner) departs from the configuration indicated by the others. An interesting feature is that this departure is more clearly indicated in the beta type II plot of Figure A_2 than in the "equivalent" beta type I plot of Figure A_1. Transforming to a distribution with infinite tails from one with finite tails may increase the "sensitivity" of the probability plot to departures in the tails.

Figures A_3 and A_4 show beta types I and II plots for the same data using the maximum likelihood estimates when $R = 1$. In spite of the estimates themselves being very different from those when $R = 0.5$, the configurations of the probability plots are not very different. Again the largest point is indicated



Data From Talker Identification Problem
Beta One Plot Alpha Estimate = 10.479 Beta Estimate = 19.378
Number of Points on Graph = 20 Sample Size = 20 Parameters were Estimated with 10 Smallest Order Statistics

Figure A_1



Data from Talker Identification Problem
Beta Two Plot Alpha Estimate = 10.479 Beta Estimate = 19.376
Number of Points on Graph = 20 Sample Size = 20 Parameters were
Estimated with 10 Smallest Order Statistics

Figure A_2

as being overly large although, as one would expect, the indication of the aberrant point is not as clear in Figure A_4 as it is in Figure A_2.

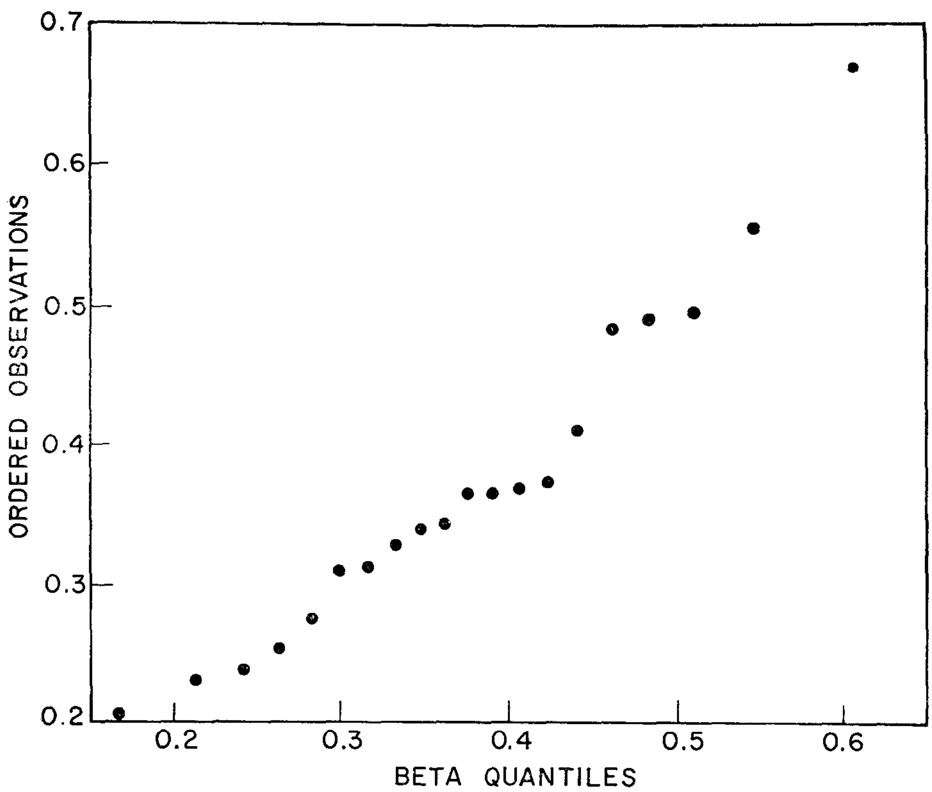
6. Concluding Remarks

Analytical methods based on the approach of Halperin (1952) could be used to study the large sample properties of the maximum likelihood estimates obtained by the methods described in the preceding sections. Useful small sample indications concerning bias, variance and covariance of these estimates could also be obtained from a systematic Monte Carlo study using computer-generated random samples from beta distributions. It would also be of interest, in such a study, to assess the maximum likelihood estimates in terms of some criterion of linearity of the configuration of the sample in the beta probability plot whose axis of quantiles is determined by using the estimates of α and β for that sample. Such a Monte Carlo study of various statistical properties of estimators obtained by different techniques for a wide class of distributions, including the gamma and beta distributions, was envisaged and initiated by the present authors but has not been completed. The partial results, as typified by the first example of Section 5, indicated that the maximum likelihood estimates obtained by the methods described in the present paper have reasonable statistical properties which justify publication of the methods involved even though definitive information on the properties of the estimates still needs to be sought and provided.

The techniques of the present paper are available for routine use on a high-speed computer and Tables 2 and 3 as well as the probability plots show typical output from the programs.

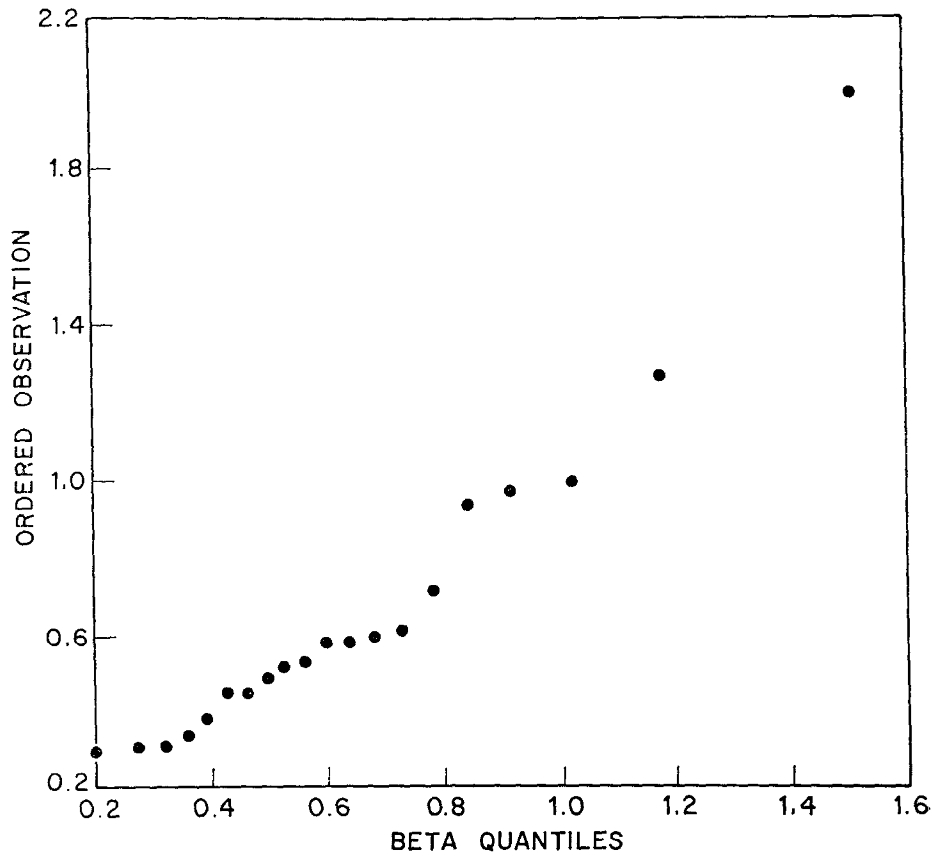
Appendix on Numerical Methods

(1) The digamma function, $\Psi(x) = \Gamma'(x)/\Gamma(x)$, is evaluated using the relationship [See Wilk et al. (1962b)]



Data from Talker Identification Problem
Beta One Plot Alpha Estimate = 6.543 Beta Estimate = 11.052
Number of Points on Graph = 20 Sample Size = 20 Parameters were Estimated with 20 Smallest Order Statistics

Figure A_3



Data from Talker Identification Problem
Beta Two Plot Alpha Estimate = 6.543 Beta Estimate = 11.052
Number of Points on Graph = 20 Sample size = 20 Parameters were
Estimated with 20 Smallest Order Statistics

Figure A_4

$$\Psi(x) \approx \begin{cases} \frac{1}{2}\ln\left[(10 + x_f)(11 + x_f)\right] + \dfrac{1}{6(10 + x_f)(11 + x_f)} \\ \qquad - \left[\dfrac{1}{x} + \dfrac{1}{x+1} + \cdots + \dfrac{1}{10 + x_f}\right], & \text{if } 0 < x < 11, \\ \frac{1}{2}\ln\left[x(x-1)\right] + \dfrac{1}{6x(x-1)}, & \text{if } x \geq 11, \end{cases}$$

where x_f is the fractional part of x. As stated in Wilk et al. (1962b), the absolute error in the approximation is less than 3×10^{-6} for moderate and large $x(\geq 3)$. For $x = 1.5$, the absolute error is less than 8×10^{-6}.

(2) The incomplete beta function,

$$I(x; \alpha, \beta) = \int_x^1 t^{\alpha-1}(1 - t)^{\beta-1}\, dt,$$

is evaluated using the first few terms of a modified Laplacian expansion (Molina (1932), Gnanadesikan, Hughes and Pinkham (1966)),

$$I(x;\alpha,\beta) \approx \sum_{j=0}^{6} \frac{A_j}{j!} \frac{1}{N^{\beta+j}} z^{\beta+j} D(\beta+j, z),$$

where

$$N = \alpha + \frac{\beta}{2} - \frac{1}{2}, \qquad z = -N \ln x,$$

$$A_0 = 1,$$

$$A_1 = A_3 = A_5 = 0,$$

$$A_2 = \frac{\beta - 1}{12},$$

$$A_4 = \frac{(\beta-1)(5\beta-7)}{240},$$

$$A_6 = \frac{(\beta-1)(35\beta^2 - 112\beta + 93)}{4032},$$

and

$$D(a, b) = \int_0^1 t^{a-1} e^{-bt}\, dt, \qquad a > 0, \qquad b > 0.$$

[NOTE: The integral, $D(a, b)$, may be evaluated using the computational procedure described in Wilk et al. (1962a).] The above expression for $I(x; \alpha, \beta)$ is used when $N > 0$, i.e., $2\alpha + \beta > 1$. For values of N between $-\frac{1}{2}$ and 0, i.e., $0 < 2\alpha + \beta \leq 1$, the recurrence relation,

$$I(x;\alpha,\beta) = I(x;\alpha+1,\beta) + I(x;\alpha,\beta+1),$$

may be used to obtain $I(x; \alpha, \beta)$.

(3) The partial derivatives, $I_1(x; \alpha, \beta)$ and $I_2(x; \alpha, \beta)$, are computed using procedures described in Gnanadesikan, Hughes and Pinkham (1966). The first seven terms of the modified Laplacian expansion for $I_1(x; \alpha, \beta)$ yield

$$I_1(x;\alpha,\beta) = \frac{\partial}{\partial\alpha} I(x;\alpha,\beta) \approx -\sum_{j=0}^{6} \frac{A_j}{j!} \frac{1}{N^{\beta+j+1}} z^{\beta+j+1} D(\beta+j+1, z),$$

where N, z, D and the coefficients A_j are the same as defined above.

For small values of x, α and β the error in using the above series is worse than for larger values. For instance, when $x = 0.01$, $\alpha = 1$ and $\beta = 3$, the absolute error is 1×10^{-3} while the relative error is approximately 0.2%.

The above expression for computing I_1 is used when $N > 0$, and, if $-\frac{1}{2} < N \leq 0$, the recurrence relation,

$$I_1(x;\alpha,\beta) = I_1(x;\alpha+1,\beta) + I_1(x;\alpha,\beta+1),$$

The integral, $I_2(x; \alpha, \beta)$, is obtained using the relationship

$$I_2(x;\alpha,\beta) = \frac{\partial}{\partial\beta} I(x;\alpha,\beta) = B(\alpha,\beta)[\Psi(\beta) - \Psi(\alpha+\beta)] - I_1(1-x;\beta,\alpha).$$

ACKNOWLEDGMENT

It is a pleasure to thank M. B. Wilk for his comments on an earlier draft.

REFERENCES

1. ANSCOMBE, F. J., 1961. Testing to establish a high degree of safety or reliability. *Bull. Int. Statist. Inst.*, *39*, 73–89.
2. BECKER, MRS. M. H., GNANADESIKAN, R., MATHEWS, M. V., PINKHAM, R. S., PRUZANSKY, MISS S., and WILK, M. B., 1965. Comparison of some statistical distance measures for talker identification. Unpublished Bell Telephone Laboratories Memo.
3. DAVID, F. N., and JOHNSON, N. L., 1948. The probability integral transformation when parameters are estimated from the sample. *Biometrika, 35*, 182–90.
4. GNANADESIKAN, R., HUGHES, MRS. L. P. and PINKHAM, R. S., 1966. Numerical evaluation of partial derivatives of the incomplete beta integral. Unpublished Bell Telephone Laboratories Memo.
5. HALPERIN, MAX, 1952. Maximum likelihood estimation in truncated samples. *Ann. Math. Statist. 23*, 226–38.
6. MOLINA, E. C., 1932. An expansion for Laplacian integrals in terms of incomplete gamma functions, and some applications. *Bell Syst. Tech. J. 11*, 563–75.
7. WILK, M. B. and GNANADESIKAN, R., 1961. Graphical analysis of multiresponse experimental data using ordered distances. *Proc. Nat. Acad. Sci. 47*, 1209–12.
8. WILK, M. B. and GNANADESIKAN, R., 1964. Graphical methods for internal comparisons in multiresponse experiments. *Ann. Math. Statist. 35*, 613–31.
9. WILK, M. B., GNANADESIKAN, R. and FREENY, MRS. A. E., 1963. Estimation of error variance from smallest ordered contrasts. *J. Amer. Statist. Assoc., 58*, 152–60.
10. WILK, M. B., GNANADESIKAN, R. and HUYETT, MISS M. J., 1962a. Probability plots for the gamma distribution. *Technometrics, 4*, 1–20.
11. WILK, M. B., GNANADESIKAN, R. and HUYETT, MARILYN J., 1962b. Estimation of parameters of the gamma distribution using order statistics. *Biometrika, 49*, 525–45.
12. WILK, M. B., GNANADESIKAN, R. and LAUH, ELIZABETH, 1966. Scale parameter estimation from the order statistics of unequal gamma components. *Ann. Math. Statist., 37*, 152–76.
13. WISE, M. E., 1950. The incomplete beta function as a contour integral and a quickly converging series for its inverse. *Biometrika, 37*, 208–18.