THEMIS SIGNAL ANALYSIS STATISTICS RESEARCH PROGRAM

SAMPLE SIZE REQUIRED TO ESTIMATE A PARAMETER IN THE POWER FUNCTION DISTRIBUTION

by

R. E. Kromer and C. H. Kapadia

Technical Report No. 13
Department of Statistics THEMIS Contract

September 25, 1968

Research sponsored by the Office of Naval Research
Contract N00014-68-A-0515
Project NR 042-260

DEPARTMENT OF STATISTICS
Southern Methodist University

Sample Size Required to Estimate a Parameter in the Power Function Distribution

by

R. E. Kromer and C. H. Kapadia*
Texas Instruments, Southern Methodist University and
Graduate Research Center of the Southwest

The purpose of this paper is to establish a two step sampling procedure for estimating the parameter θ of the power function distribution to within given d units of its true value with a given probability $1 - \alpha$; $(0 < \alpha < 1)$. The density of the power function distribution is a function of two parameters, the second of which k is assumed known. It is demonstrated that an exact solution for all values of θ does not exist based on the maximum likelihood estimator. Given a preliminary sample size m, tables and formulas are presented by which one may establish the size n of the second sample such that $P(|y_n - \theta| < d) > 1 - \alpha$ is true, where y_n is the largest observation in the second sample. The method used in deriving the results of this paper is similar to that given by Graybill and Connell [2(hereafter referred to as GC)] and since the power function density reduces to the uniform density when k = 0, their results can be derived from the formulas given here. Also a table of comparisons between the expected second sample size in this paper and two other solutions is given.

1. Introduction

Let $f(u) = (k + 1)\,\theta^{-(k+1)}\,u^k$, $0 < u \le \theta$, $\theta > 0$ and zero elsewhere be a power function density. The maximum likelihood estimator for θ based on a sample of size n is the largest value y_n. If $\hat{\theta} = y_n$ and $\hat{\theta}^* = \{n(k+1) + 1\}y_n/n(k+1)$ are respectively biased and unbiased estimator of θ then it can be shown that for at least one set of values for a preliminary sample size m and given k

$$P(|\hat{\theta} - \theta| < d) > P(|\hat{\theta}^* - \theta| < d)$$

is true for certain values of d, while for others, the reverse is true. Thus, since neither $\hat{\theta}^*$ nor $\hat{\theta}$ is uniformly better for all values of d, $\hat{\theta}$ will be used in this paper. Hence the problem is to determine n, the size of a sample from this distribution, such that the estimator $\hat{\theta}$ based on the sample will have probability of at least $1 - \alpha$ of being

---

* Research supported in part by NIH Training Grant GM-951, and revised under ONR Contract N00014-68-A-0515.

within d units of θ, where d and $1 - \alpha$ are specified in advance, $0 < \alpha < 1$, symbolically,

$$P\{|y_n - \theta| < d\} > 1 - \alpha \,. \tag{1}$$

The solution will be a two step procedure where n will be determined on the basis of a preliminary sample of size m.

2. Solution

If $a = d/\theta$ and E_t denotes expectation with respect to the random variable t, then (1) can be rewritten as

$$\begin{aligned} P[|y_n - \theta| < d] &= 1 - E_n[P(1 - a < (Y_{n/\theta}) < 1|n)] \\ &= 1 - E_n[(1 - a)^{(k+1)n}] \,. \end{aligned} \tag{2}$$

When $d \geq \theta$, $P(|y_n - \theta| < d) = 1$; hence this case will not be considered.

Consider a preliminary sample of size m taken from f(u) and let x be the largest observation. The second sample size must be a function of x, i.e., $n = t(x)$. If the function t(x) has all the properties given in equation (3) of GC, then it can be shown that Theorem 1 given by GC is also true here for the equation

$$E_n[(1 - a)^{(k+1)n}] = \alpha \,.$$

If θ were known, then from equations (1) and (2),

$$n = (\log \alpha)/\{(k+1) \log (1-a)\} \,,\ 0 < a < 1 \,,\ 0 < \alpha < 1 \,,\ k \geq 0 \,. \tag{3}$$

Since θ is not known, then following the procedure and conditions given in GC [p. 553], the sample size n becomes

$$n = \begin{cases} 1 & ,\ 0 \leq x \leq bd/(1-\alpha^{1/k+1}) \,, \\ (\log \alpha)/\{(k+1) \log (1-bd/x)\}, & bd/(1-\alpha^{1/k+1}) \leq x \leq \theta \,, \end{cases} \tag{4}$$

where b is as given in GC [p.553] and the function in (4) has all the properties given in GC [p.553].

If $z = x/\theta$, then the density function for z is

$$h(z) = \begin{cases} m(k+1)\, z^{m(k+1)-1}, & 0 \le z \le 1, \quad k \ge 0, \\ 0, & \text{elsewhere.} \end{cases}$$

Let

$$f_a(z) = \begin{cases} (1-a)^{k+1} & , \ 0 \le z \le ba/(1-\alpha^{1/k+1}) \\ (1-a)^{(\log \alpha)/\{\log [1-(ba/z)]\}} & , \ ba(1-\alpha^{1/k+1}) \le z \le 1 . \end{cases} \tag{5}$$

For the remainder of this paper, n will denote the size of the random sample from f(u), where n is given in equation (4). Therefore

$$E_n\left[(1-a)^{n(k+1)}\right] = \int_0^1 f_a(z)\, h(z)\, dz.$$

If $g_0(z)$ is any function which has all the properties given in equation (8) of GC, then

$$E_n\left[(1-a)^{n(k+1)}\right] = \int_0^1 f_a(z)\, h(z)\, dz \le \alpha, \quad 0 < a < 1 .$$

Theorem: The smallest function having properties given in equation (8) of GC is the function g(z), where

$$g(z) = \begin{cases} [1 - \{1 - \alpha^{1/(k+1)}\}(z/b)]^{k+1} & 0 \le z \le b \\ \alpha^{z/b} & b \le z \le 1 . \end{cases} \tag{6}$$

The proof of the theorem runs parallel to that of Theorem 2 in GC and hence the lemmas which are useful to prove the same will be given.

Lemma 1:

For two real numbers existing in the intervals $0 \le q \le 1$ and $0 \le r \le 1$, such that $q \ge 1 - r$, the following inequality holds:

$$q^{1/(1-q)} \ge (1-r)^{1/r} .$$

Lemma 2:

$\hat{f}_a(z)$ is convex in the interval $ba/\left[1 - \alpha^{1/(k+1)}\right] \le z \le b$ where

$$\hat{f}_a(z) = (1 - a)^{[\log \alpha^{1/(k+1)}]/\{\log [1 - (ba/z)]\}} .$$

Lemma 3:

From equation (5), it follows that

$$\lim_{a \to 0} f_a(z) = \alpha^{z/b} , \quad 0 < z \le 1 .$$

Lemma 4:

$f_a(z)$ is monotonic decreasing in a for

$$z \ge \max \left[b, (ba)/\left(1 - \alpha^{1/(k+1)}\right)\right].$$

3. Summary

From the above theorem, it follows that

$$E_n[(1-a)^{n(k+1)}] = \int_0^1 f(z)\, h(z)\, dz$$

$$\le \int_0^1 g(z)\, h(z)\, dz$$

$$= m(k+1)\, b^{m(k+1)} \sum_{r=0}^{k+1} \binom{k+1}{r} [\alpha^{1/k+1} - 1]^r / \{r + m(k+1)\}$$

$$+ \; m(k+1) \int_b^1 \alpha^{z/b}\, z^{m(k+1) - 1}\, dz . \qquad (7)$$

If equation (7) is set equal to α and solved implicitly for b for each m, α and k, it is assured that the use of this b in equation (4) will give a second sample size n such that (1) is satisfied. Tables 1 to 6 give values of b such that (7) is equal to α.

4. Sample Size Tables

To determine the desired second sample size n, first find the value of b in Tables 2 and 3 depending on the choice for m (the preliminary sample size) $1 - \alpha$ and k. Then find x, the largest observation in the preliminary sample, determine d and find n according to

$$n = \begin{cases} 1 & , \ 0 \le x \le (bd)/(1-\alpha^{1/k+1}) \ , \\ [\log \alpha]/\{(k+1) \log (1 - bd/x)\} & , \ bd/(1-\alpha^{1/k+1}) \le x \le \theta . \end{cases}$$

The use of this n for the second sample size guarantees that

$$P[|y_n - \theta| < d] > 1 - \alpha ,$$

where y_n is the largest observation in the second sample. If it occurs that x is greater than y_n, then x satisfies equation (1) and is superior to y_n as an estimator of θ.

The contents of Tables 2 and 3 are similar to those given in GC except the sample size, determined by the use of Tchebycheff's inequality [1], may be shown to be equal to

$$n = [1/(k+1)]\left\{1 + [m(k+1)\ x^2]/[\{m(k+1) - 2\}\ \alpha d^2]\right\}^{1/2} - 1/(k+1) .$$

The conclusions derived from these tables are also exactly the same as those given in GC.

Table I

Values of b for Finding Sample Size n

1-α	.90					
m \ k	0	1	2	3	5	10
2	.4420	.7101	.8110	.8613	.9101	.9525
5	.7282	.8772	.9229	.9441	.9641	.9811
10	.8534	.9378	.9614	.9721	.9821	.9905
15	.8996	.9584	.9742	.9814	.9880	.9937
20	.9237	.9688	.9807	.9860	.9910	.9953
25	.9385	.9750	.9845	.9888	.9928	.9962
30	.9484	.9792	.9871	.9907	.9940	.9968
40	.9610	.9844	.9903	.9930	.9955	.9976
50	.9687	.9875	.9923	.9944	.9964	.9981
75	.9790	.9917	.9949	.9963	.9976	.9987
100	.9842	.9937	.9961	.9972	.9982	.9991

1-α	.95					
m \ k	0	1	2	3	5	10
2	.3489	.6690	.7909	.8499	.9052	.9511
5	.6818	.8634	.9170	.9410	.9628	.9807
10	.8292	.9318	.9589	.9708	.9815	.9904
15	.8833	.9547	.9728	.9806	.9877	.9936
20	.9113	.9661	.9796	.9855	.9908	.9952
25	.9285	.9729	.9837	.9884	.9926	.9962
30	.9401	.9774	.9864	.9903	.9939	.9968
40	.9548	.9831	.9898	.9928	.9954	.9976
50	.9637	.9865	.9919	.9942	.9963	.9981
75	.9757	.9910	.9946	.9961	.9976	.9987
100	.9817	.9933	.9959	.9971	.9982	.9990

1-α	.99					
m \ k	0	1	2	3	5	10
2	.1702	.5422	.7274	.8148	.8909	.9475
5	.5492	.8200	.8995	.9321	.9593	.9798
10	.7585	.9135	.9520	.9673	.9802	.9900
15	.8356	.9436	.9686	.9785	.9869	.9934
20	.8755	.9582	.9767	.9840	.9902	.9950
25	.8997	.9668	.9815	.9873	.9922	.9960
30	.9161	.9725	.9846	.9894	.9935	.9967
40	.9368	.9796	.9885	.9921	.9951	.9975
50	.9492	.9837	.9909	.9937	.9961	.9980
75	.9660	.9892	.9939	.9958	.9974	.9987
100	.9745	.9919	.9955	.9969	.9981	.9990

TABLE 2.—Expected second sample sizes for k = 1

a	m \ $1-\alpha$	.90 Ideal	.90 $E_1(n)$	.90 $E_2(n)$	.95 Ideal	.95 $E_1(n)$	.95 $E_2(n)$	.99 Ideal	.99 $E_1(n)$	.99 $E_2(n)$
.05	5	22.4	23.3	31.6	29.2	30.8	45.0	44.9	49.9	101.1
.05	20	22.4	22.6	31.2	29.2	29.5	44.3	44.9	45.7	99.6
.05	50	22.4	22.5	31.1	29.2	29.3	44.2	44.9	45.2	99.5
.25	5	4.0	4.2	5.9	5.2	5.5	8.6	8.0	9.0	19.8
.25	20	4.0	4.0	5.9	5.2	5.3	8.5	8.0	8.2	19.5
.25	50	4.0	4.0	5.8	5.2	5.2	8.5	8.0	8.1	19.5
.60	5	1.3	1.3	2.2	1.6	1.8	3.3	2.5	3.0	8.0
.60	20	1.3	1.3	2.2	1.6	1.7	3.3	2.5	2.6	7.9
.60	50	1.3	1.3	2.2	1.6	1.6	3.3	2.5	2.5	7.8

TABLE 3. —Expected second sample sizes for k = 3

a	m \ $1-\alpha$	.90 Ideal	.90 $E_1(n)$	.90 $E_2(n)$	.95 Ideal	.95 $E_1(n)$	.95 $E_2(n)$	.99 Ideal	.99 $E_1(n)$	.99 $E_2(n)$
.05	5	11.2	11.3	15.6	14.6	14.8	22.2	22.4	22.9	49.9
.05	20	11.2	11.2	15.6	14.6	14.6	22.1	22.4	22.5	49.8
.05	50	11.2	11.2	15.6	14.6	14.6	22.1	22.4	22.5	49.8
.25	5	2.0	2.0	2.9	2.6	2.6	4.2	4.0	4.1	9.8
.25	20	2.0	2.0	2.9	2.6	2.6	4.2	4.0	4.0	9.8
.25	50	2.0	2.0	2.9	2.6	2.6	4.2	4.0	4.0	9.8
.60	5	.6	1.0	1.1	.8	1.0	1.6	1.3	1.3	3.9
.60	20	.6	1.0	1.1	.8	1.0	1.6	1.3	1.3	3.9
.60	50	.6	1.0	1.1	.8	1.0	1.6	1.3	1.3	3.9

[1] Birnbaum, A., and Healy, W. C., "Estimates with prescribed variance based on two stage sampling" Annals of Mathematical Statistics, 31 (1960), 662-76.

[2] Graybill, Franklin A., and Connell, Terrance L., "Sample size required to estimate the parameter in the uniform density within d units of the true value" Journal of American Statistical Association, 59 (1964) 550-56.