Ann. Inst. Statist. Math.
39 (1987), Part A, 627-635

ON THE LOGISTIC MIDRANGE

E. OLUSEGUN GEORGE AND CECIL C. ROUSSEAU

(Received Sept. 5, 1985; revised Jan. 8, 1986)

Summary

It is well-known that for a large family of distributions, the sample midrange is asymptotically logistic. In this article, the logistic midrange is closely examined. Its distribution function is derived using Dixon's formula (Bailey (1935, *Generalized Hypergeometric Series*, Cambridge University Press, p. 13)) for the generalized hypergeometric function with unit argument, together with appropriate techniques for the inversion of (bilateral) Laplace transforms. Several relationships in distribution are established between the midrange and sample median of the logistic and Laplace random variables. Possible applications in testing for outliers are also discussed.

1. Introduction

The logistic distribution, which in standard form has distribution function

$$F_0(x)=[1+\exp(-x)]^{-1}, \qquad -\infty<x<\infty, \tag{1.1}$$

has long been applied in a variety of statistical studies: Pearl and Reed [10], Verhulst [12] and several authors used it in the study of population growth; Amemiya [1], Berkson [3] and others employed it in the analysis of bioassay data; Cox [4] and in several later articles, used it for analyzing data from binary experiments; Plackett [11] considered its use in problems involving censored data; Gumbel [7] showed that it is the limiting distribution of the standardized midrange and the extremal quotient of a wide family of distributions.

Among the family of distributions considered by Gumbel is the logistic itself. Gumbel's result shows that the asymptotic distribution of the standardized logistic midrange is itself logistic. In this article, we closely examine the logistic midrange. We obtain the characteristic

Key words and phrases: Characteristic functions, sample median, logistic distribution, Laplace distribution.

function and the exact distribution of the logistic midrange and in the process we exhibit an interesting functional duality displayed by the logistic midrange depending on whether the sample size n is odd or even. This duality carries over to the distribution of the non-standardized midrange, producing a function of the logistic distribution when n is odd and a function of the convolution of two logistic distributions when n is even. In addition, we exhibit an interplay between the logistic distribution, the double exponential distribution and the distribution of the midrange. In Section 2, the characteristic function is used to compute the exact distribution function of the logistic midrange. A discussion of the relationship to the double exponential, some of the consequences with regard to ordered statistics from the logistic distribution and the asymptotic properties of the logistic midrange are given in Section 3. A possible use of the midrange to detect outliers is outlined in Section 4.

2. The exact distribution of the logistic midrange

Let $X_1, \cdots, X_n$ be a random sample from a logistic distribution with distribution function given by (1.1). Let $X_{(1)} \leqq X_{(2)} \leqq \cdots \leqq X_{(n)}$ be the ordered statistics and let

$$M_n = (X_{(1)} + X_{(n)})/2 \tag{2.1}$$

be the midrange.

THEOREM 2.1. *The characteristic function $\phi_n(t)$ of M_n is given by*

$$\phi_n(t) = \begin{cases} \prod_{j=1}^{p-1} (1+t^2/4j^2)[(\pi\, it/2)/\sin(\pi\, it/2)]^2 , & \text{for } n=2p , \\ \prod_{j=1}^{p} [1+t^2/(2j-1)^2][\pi\, it/\sin \pi\, it] , & \text{for } n=2p+1 . \end{cases} \tag{2.2}$$

PROOF. The characteristic function of M_n is given by

$$\phi_n(t) = n(n-1) \int_{-\infty}^{\infty} \int_{-\infty}^{y} e^{it(x+y)/2} [F_0(y) - F_0(x)]^{n-2} f_0(x) f_0(y) dx dy .$$

Using the transformations $u = F_0(x)$, $v = F_0(y)$, we get

$$\phi_n(t) = n(n-1) \int_0^1 \int_0^v \{uv/[(1-u)(1-v)]\}^{it/2} (v-u)^{n-2} du dv . \tag{2.3}$$

Expanding $(1-v)^{it}$ and integrating term by term, we find

$$\phi_n(t) = n(n-1) \sum_{k=0}^{\infty} (it/2)_k B(k+n+it, 1-it/2) B(k+1+it/2, n-1)/k! ,$$

where $B(x, y)$ denotes the beta function, and $(x)_k = \Gamma(k+x)/\Gamma(x)$. Using $B(k+x, y) = B(x, y)(x)_k/(x+y)_k$, we express $\psi_n(t)$ in terms of a well-poised hypergeometric function ${}_3F_2$, *viz.*

$$\psi_n(t) = G_n(it)\,{}_3F_2\begin{bmatrix} n+it, & it/2, & 1+it/2\,; \\ n+1+it/2, & & n+it/2 \end{bmatrix},$$

where

$$G_n(it) = n(n-1)B(n+it, 1-it/2)B(1+it/2, n-1)\,.$$

Hence, from Dixon's formula (see Bailey [2], p. 13) we get

$$\psi_n(t) = \frac{\Gamma(1+it/2)\Gamma(1-it/2)\Gamma\{(n+it)/2\}\Gamma\{(n-it)/2\}}{\Gamma^2(n/2)}\,. \tag{2.4}$$

For further details, see George and Rousseau [6]. Thus if $n=2p$, (2.4) reduces to

$$\psi_{2p}(t) = [\Gamma(1+it/2)\Gamma(1-it/2)]^2 \prod_{j=1}^{p-1} (1+t^2/4j^2) \tag{2.5}$$

$$= [(\pi\, it/2)/\sin(\pi\, it/2)]^2 \prod_{j=1}^{p-1} (1+t^2/4j^2)\,, \tag{2.6}$$

and when $n=2p+1$, (2.4) reduces to

$$\psi_{2p+1}(t) = \Gamma(1+it/2)\Gamma(1-it/2)\sec(\pi\, it/2) \prod_{j=1}^{p} [1+t^2/(2j-1)^2]\,, \tag{2.7}$$

$$= (\pi\, it/2)\operatorname{cosec}(\pi\, it/2)\sec(\pi\, it/2) \prod_{j=1}^{p} [1+t^2/(2j-1)^2]$$

$$= [\pi\, it/\sin \pi\, it] \prod_{j=1}^{p} [1+t^2/(2j-1)^2]\,. \tag{2.8}$$

THEOREM 2.2. *The density function of the midrange from a logistic distribution is a Polya frequency function.*

PROOF. It is well-known that

$$\frac{\sin(i\pi t)}{i\pi t} = \prod_{j=1}^{\infty} (1+t^2/j^2)\,.$$

Hence, $\rho_n(t) = 1/\psi_n(t)$ is an entire function given by

$$\rho_n(t) = \prod_{k=1}^{\infty} (1+t^2/4k^2) \prod_{j=p}^{\infty} (1+t^2/4j^2)\,, \tag{2.9}$$

for $n=2p$, and by

$$\rho_n(t) = \prod_{k=1}^{p} (1+t^2/4k^2) \prod_{j=n}^{\infty} (1+t^2/j^2)\,, \tag{2.10}$$

when $n=2p+1$. It follows immediately from Theorem 3.2 a, p. 345 of Karlin [8] that the density function of M_n is a Polya frequency function.

In order to express the distribution function of M_n in closed form, it is helpful to introduce what we shall call the *logistic polynomials*. Define the sequence of logistic polynomials $\{P_n(x)\}$ by

$$P_0(x)=1\,, \qquad P_n(x)=D\{x(1-x)P_{n-1}(x)\}\,, \qquad n\geqq 1\,, \tag{2.11}$$

where $D=d/dx$. Thus, the first few logistic polynomials are $P_0(x)=1$, $P_1(x)=1-2x$, $P_2(x)=1-6x+6x^2$, $P_3(x)=1-14x+36x^2-24x^3$.

The above are called logistic polynomials in view of the following result:

LEMMA 2.1. *If F satisfies the logistic differential equation, $DF=F(1-F)$, then its higher derivatives are given by*

$$D^nF=F(1-F)P_{n-1}(F)\,, \qquad n=1, 2, 3, \cdots\,. \tag{2.12}$$

PROOF. The proof is a simple consequence of (2.11) and the chain rule.

It is easy to find an explicit formula for P_n. Writing

$$P_n(x)=\sum_{k=0}^{n}(-1)^k c(n, k)x^k\,, \tag{2.13}$$

the differential recurrence (2.11) yields the recurrence formula

$$c(n, k)=(k+1)[c(n-1, k)+c(n-1, k-1)]\,. \tag{2.14}$$

Equation (2.14), subject to the condition $P_0(x)=1$, is easily solved, with the result

$$c(n, k)=(k+1)!\,S(n+1, k+1)\,, \tag{2.15}$$

where $S(\cdot\,, \cdot)$ denotes a Stirling number of the second kind.

In the following theorem, we shall apply inversion methods to the results obtained in Theorem 2.1 in order to calculate the exact distribution function of M_n. In doing so, we make the substitution $t=is$ and express the previous result (2.2) in the language of (bilateral) Laplace transforms.

THEOREM 2.3. *Let $X_1, \cdots, X_n$ be a random sample from the logistic distribution with distribution function $F_0(x)=[1+\exp(-x)]^{-1}$, $-\infty<x<\infty$. Let $G_n(x)$ denote the distribution function of the midrange $M_n=(X_{(1)}+X_{(n)})/2$. If $a_1, a_2, \cdots, a_p$ and $b_1, b_2, \cdots, b_{p-1}$ are defined by*

$$Q_p(x)=\prod_{j=1}^{p}\left\{1-\frac{x^2}{(2j-1)^2}\right\}=1+\sum_{k=1}^{p}(-1)^k a_k x^{2k} \tag{2.16}$$

and

$$R_p(x)=\prod_{j=1}^{p-1}\left\{1-\frac{x^2}{j^2}\right\}=1+\sum_{k=1}^{p-1}(-1)^k b_k x^{2k} \tag{2.17}$$

respectively, then

$$G_{2p+1}(x)=F_0(x)+\sum_{k=1}^{p}(-1)^k a_k D^{2k}F_0(x) \tag{2.18}$$

and

$$G_{2p}(x/2)=D(xH_0(x))+\sum_{k=1}^{p-1}(-1)^k b_k D^{2k+1}(xH_0(x)) \ , \tag{2.19}$$

where $F_0(x)=(1+\exp(-x))^{-1}$, $H_0(x)=(1-\exp(-x))^{-1}$ *and* $D=d/dx$. *In terms of the logistic polynomials, the requisite derivatives are given by*

$$\begin{aligned} &D^rF_0=F_0(1-F_0)P_{r-1}(F_0)\ , \\ &D(xH_0)=\begin{cases} \dfrac{1-(1+x)e^{-x}}{(1-e^{-x})^2}\ , & x\neq 0\ , \\ 1/2\ , & x=0\ , \end{cases} \end{aligned} \tag{2.20}$$

and for $r\geqq 2$,

$$D^r(xH_0)=\begin{cases} H_0(1-H_0)[xP_{r-1}(H_0)+rP_{r-2}(H_0)]\ , & x\neq 0\ , \\ (-1)^rB_r\ , & x=0\ , \end{cases} \tag{2.21}$$

where B_r *is the r-th Bernoulli number.*

PROOF. From (2.2), we have the (bilateral) Laplace-Stieltjes transform

$$\int_{-\infty}^{\infty} e^{-sx}dG_{2p+1}(x)=Q_p(x)\pi s\ \mathrm{cosec}\,(\pi s)\ . \tag{2.22}$$

To invert this, we simply use well-known properties of the Laplace transform along with the fact that $\pi\,\mathrm{cosec}\,(\pi s)$ is the transform of F_0. We thus obtain (2.18). Equation (2.20) is simply a re-statement of Lemma 2.1.

Again from (2.2), we know that

$$\int_{-\infty}^{\infty} e^{-sx}dG_{2p}(x)=R_p(s/2)(\pi s/2)^2\ \mathrm{cosec}^2\,(\pi s/2)\ . \tag{2.23}$$

Using an obvious scale change and the fact that $\{\pi\,\mathrm{cosec}\,(\pi s)\}^2$ is the Laplace transform of $xH_0(x)$, we obtain (2.19). Since H_0 is a solution of the logistic equation, we know that

$$D^r H_0 = H_0(1-H_0)P_{r-1}(H_0) . \tag{2.24}$$

Now (2.21) follows from $D^r(xH_0)=xD^rH_0+rD^{r-1}H_0$, which is a simple consequence of Leibnitz rule.

Remark. The individual terms of the first line of (2.21) on the right hand side are badly behaved near $x=0$. This is the price we pay for splitting H_0 from xH_0 in order to be able to write $D^r(xH_0)$ as a finite sum (in terms of logistic polynomials). For numerical calculation of $D^r(xH_0)$ near $x=0$, one should perhaps abandon (2.21) in favor of the well-known infinite series

$$D^r(xH_0)=\sum_{k=0}^{\infty}(-1)^{r+k}B_{r+k}x^k/k! , \qquad |x|<2\pi , \tag{2.25}$$

where B_j denotes the j-th Bernoulli number.

The coefficients $a_1,\cdots,a_p$ and $b_1,\cdots,b_{p-1}$ in Q_p and R_p respectively are given by

$$a_k=(-1)^{p-k}2^{4p}(p!)^2\sum_{q=2k}^{2p}(-1/2)^q S(2p, q)C(q, 2k)(2p-1)^{q-2k}/[(2p)!]^2 ,$$

and

$$b_k=(-1)^k\sum_{q+r=2k}(-1)^q S(p+1, q+1)S(p+1, r+1)/(p!)^2 ,$$

where $C(\cdot, \cdot)$ denotes a binomial coefficient and $S(\cdot, \cdot)$ denotes a Stirling number of the first kind (see Knuth [9]).

Asymptotic formulas for G_n are readily obtained using the techniques of Theorem 2.3 together with the well-known asymptotics of the gamma function. From the basic result (2.4), we have the fact that $G_n(x/2)$ is the inverse Laplace transform of $\pi\operatorname{cosec}(\pi s)\Gamma(n/2+s)\cdot\Gamma(n/2-s)/\Gamma^2(n/2)$. Now the asymptotic approximation

$$\frac{\Gamma(n/2+s)\Gamma(n/2-s)}{\Gamma^2(n/2)}=1+2s^2/n+2s^2(s^2+1)/n^2+O(1/n^3) ,$$

yields

$$G_n(x/2)=y+y(1-y)[2P_1(y)/n+2\{P_3(y)+P_1(y)\}/n^2+O(1/n^3)] , \tag{2.26}$$

where $y=F_0(x)$.

3. Some properties of the logistic midrange

Let $X_1, X_2,\cdots, X_n$ be independent logistic r.v.s. with distribution function $F_0(x)=(1+e^{-x})^{-1}$, and let $Z_1,\cdots, Z_n$ be independent Laplace random variables where Z_j has parameter j, $j=1,\cdots, n$. Specifically, for

$j=1,\cdots,n$, the density of Z_j is given by

$$f_j(z)=2^{-1}je^{-j|z|}\,, \qquad -\infty<z<\infty \tag{3.1}$$

and the characteristic function of Z_j by

$$\psi_j(t)=(1+t^2/j^2)^{-1}\,, \qquad -\infty<t<\infty\,. \tag{3.2}$$

Let L denote equality in distribution. Then it is immediately clear from equation (2.2), that when $n=2p$

$$M_{2p}+\prod_{j=1}^{p} Z_{2j}\overset{L}{=}(X_1+X_2)/2\,, \tag{3.3}$$

where $M_{2p}=(X_{(1,2p)}+X_{(2p,2p)})/2$ and $X_{(1,2p)}\leqq X_{(2,2p)}\leqq\cdots\leqq X_{(2p,2p)}$ are the order statistics of $X_1,\cdots,X_{2p}$. Furthermore, the characteristic function of $X_{(1,2p)}+X_{(2p,2p)}$ is given by

$$\psi_{2p}(2t)=\prod_{j=1}^{p}(1+t^2/j^2)(\pi\, it/\sin\pi\, it)^2\,.$$

Hence, we get

$$X_{(1,2p)}+X_{(2p,2p)}+\sum_{j=1}^{p} Z_j\overset{L}{=}X_1+X_2\,. \tag{3.4}$$

Now let $n=2p+1$. The sample median $X_{(p+1)}$ has characteristic function

$$\begin{aligned}\psi_{X(p+1)}(t)&=[(2p+1)!]/[p!]^2\int_{-\infty}^{\infty}e^{itx}\{F_0(x)[1-F_0(x)]\}^p f_0(x)dx \\ &=[(2p+1)!]/[p!]^2\int_0^1[u/(1-u)]^{it}[u(1-u)]^p du \\ &=(p!)^{-2}\Gamma(p+1+it)\Gamma(p+1-it)\end{aligned} \tag{3.5}$$

$$=\prod_{j=1}^{p}(1+t^2/j^2)(\pi\, it/\sin\pi\, it)\,. \tag{3.6}$$

Hence, using the characteristic function of $2M_{2p}$ from (2.2), we have

$$\psi_{2p}(2t)=\psi_{X(p+1)}(t)(\pi\, it/\sin\pi\, it)\,.$$

Thus

$$X_{(1,2p)}+X_{(2p,2p)}\overset{L}{=}X_{(p+1,2p+1)}+Y\,, \tag{3.7}$$

where Y is an independent logistic random variable with distribution function F_0.

Still with $n=2p+1$, we have from (2.2) the obvious relation

$$M_{2p+1}+\sum_{j=1}^{p} Z_{2j-1}\overset{L}{=}X_1\,. \tag{3.8}$$

Now let $n=3$. It is easily seen from (3.4) that

$$\psi_{X_{(2)}}(t)=(1+t^2)(\pi\, it/\sin \pi\, it) .$$

Hence, we have immediately

$$(X_{(1,3)}+X_{(3,3)})/2 \stackrel{L}{=} X_{(2,3)} . \tag{3.9}$$

4. Possible applications

Although the main use of the sample midrange is in the estimation of the location parameter in a symmetric distribution, the properties of the midrange obtained above suggest a simple method of testing for the presence of an outlier. Specifically, consider a population with symmetric density f and distribution F satisfying

$$\lim_{x\to\infty} d/dx\{[1-F(x)]/f(x)\}=0 . \tag{4.1}$$

Such distributions have tails which die at an exponential rate and are thus called *exponential type*. Let $X_1, \cdots, X_n$ be an observed sample and suppose that one of these observations is suspected of being an outlier. Without loss of generality, let $X_{(n)}$ be the suspected observation. Since it is well-known that the midrange is quite sensitive to the presence of an outlier, a test using M_n as a test statistic could be considered. When F is the logistic distribution, critical points for the test of hypothesis that the sample has one outlier can be obtained using the exact distribution of M_n obtained in Section 2. More generally, for any distribution satisfying (3.13), Gumbel [7] has shown that the sample midrange is asymptotically logistic in distribution. Hence, for a large sample test, the logistic distribution can be used as the approximate distribution of M_n under the null hypothesis. For moderate sample sizes, better approximations are needed for the distribution of M_n. This is particularly so when F is the normal distribution function. In this case the exact distribution of M_n in closed form is unknown, and M_n converges very slowly to logistic in distribution (Galambos [5]). Hence, an approximation is needed here. The goodness of any such approximation can be checked using the logistic midrange for which we have already obtained the exact distribution. Constructions of various tests for outliers based on the midrange and the generalized midrange are being investigated by the authors.

MEMPHIS STATE UNIVERSITY

REFERENCES

[1] Amemiya, T. (1980). The n^{-2}-order mean squared errors of the maximum chi-square estimator, *Ann. Statist.*, **8**, 488-503.

[2] Bailey, W. N. (1935). *Generalized Hypergeometric Series*, Cambridge University Press, London.

[3] Berkson, J. (1944). Application of the logistic function to bio-assay, *J. Amer. Statist. Ass.*, **39**, 357-365.

[4] Cox, D. R. (1970). *Analysis of Binary Data*, Methuen, London.

[5] Galambos, J. (1978). *The Asymptotic Theory of Extreme Order Statistics*, Wiley, New York.

[6] George, E. O. and Rousseau, C. C. (1986). A moment generating function, Problem 85-22 in Problems and Solutions, *Soc. Industr. Appl. Math.* (Review). (to appear).

[7] Gumbel, E. J. (1944). Ranges and midranges, *Ann. Math. Statist.*, **15**, 414-422.

[8] Karlin, S. (1968). *Total Positivity*, **1**, Stanford University Press, Stanford, California.

[9] Knuth, D. E. (1973). *The Art of Computer Programming*, **1**, *Fundamental Algorithms*, Addison-Wesley Reading, Massachusetts.

[10] Pearl, R. and Reed, L. J. (1920). On the rate of growth of population of the United States since 1790 and its mathematical representation, *Proc. Nat. Acad. Sci. USA*, **6**, 275-288.

[11] Plackett, R. L. (1959). The analysis of life test data, *Technometrics*, **1**, 9-19.

[12] Verhulst, P. F. (1845). Recherches mathematiques sur la loi d'accroisement de la population, *Nouveaux memoirs de l'ac Bruxelles*, **18**, 1-38.