APPROXIMATIONS TO THE MOMENTS OF THE SAMPLE MEDIAN

By M. M. Siddiqui

Boulder Laboratories, National Bureau of Standards

0. Summary. In this paper a numerical study of Chu and Hotelling's [1] method of approximating to the moments of the sample median will be made. With an introductory outline of their method in Section 2, we will proceed to apply it to various distributions, and will evaluate the degree of accuracy that can be conveniently obtained by means of it in each particular case. The numerical results will be presented in tabular form.

1. Introduction. Chu and Hotelling [1] developed a method of approximation, using a Taylor's series expansion of the inverse function of the cumulative distribution function, for the moments of the sample median of a certain type of parent distribution. However, they did not attempt to give numerical examples, except for the normal distribution, to illustrate the rapidity of convergence of their series. The example which they worked out does not seem to be completely satisfying—an upper bound to the proportional error being 6.9×10^{-3} for samples of sizes $\geqq 101$, and 6.8×10^{-5} for samples of sizes $\geqq 501$, using the first two terms of their series to approximate the variance of the median. It seems desirable, therefore, to study the rapidity of convergence of their series by calculating the approximate variances of the medians of small samples from various special populations, using the first four or five terms. We will, in fact, use partial sums up to and including the term of relative order N^{-4}, N being the sample size.

To compare the approximate with the exact values, it is necessary to select distributions for which the exact moments, at least the first two, of the sample median for small values of N are available. Within this restricted class an attempt is made to select populations which provide some contrast and variety with respect to the range of the variate, the symmetry of the probability density function (pdf), and the "spread" of the probability mass, which three properties seem theoretically relevant to the behavior of the series. A measure of the last-mentioned property was taken to be the kurtosis, γ_2 = standardized fourth moment -3. The selected populations are the following and are ordered with respect to the range, symmetry and the value of γ_2.

(1) U-shaped (incomplete beta): finite range; symmetric; $\gamma_2 = -1.5$.
(2) Rectangular: finite range; symmetric; $\gamma_2 = -1.2$.
(3) Parabolic: finite range; symmetric; $\gamma_2 = -0.86$.
(4) Normal: infinite range; symmetric; $\gamma_2 = 0$.
(5) Double exponential: infinite range; symmetric; $\gamma_2 = 3$.
(6) Cauchy: infinite range; symmetric; moments do not exist.
(7) Exponential: semi-infinite range; skew; $\gamma_2 = 6$.

Received December 5, 1960; revised August 12, 1961.

Broadly speaking, the characteristic which is most favorable to the rapidity of convergence, is the finiteness of the range of the variate. In this case the median has finite range, and the series (4) below for $m(z)$ is convergent in the *closed* interval $-1 \leqq z \leqq 1$. If the range of the variate is infinite, $m(z)$ is unbounded in $[-1, 1]$ and the series (4) does not converge for at least one $z = -1, 1$. Symmetry of the pdf and a low value of γ_2 are also expected to be favorable. Thus it is expected that the first three populations are most favorable, the last two the least, and the remaining two fall in the middle.

We first note that, in all cases studied, the numerical results indicate that, for a fixed partial sum, the relative error $=$ | (exact-approx.)/exact | decreases monotonically with increasing N. We also note that for the rectangular population the series terminates at the first term, giving the exact value of the moment. For other populations the actual calculations, using the partial sum up to and including the term of relative order N^{-4} in each case, generally support our expectations. Thus for the relative error to be less than or equal to 0.01, the value of N must be at least as large as 1, 3, 7, 17, and 9, for the U-shaped, parabolic, normal, Cauchy and exponential distributions, respectively. The exact values of the variance of the median for $N > 5$ are not available for the double exponential distribution; however, comparisons for $N = 5$ support our general conclusions.

2. Chu and Hotelling's method. For ready reference Chu and Hotelling's method will be outlined here with slight modifications and some changes of notation. This outline is not intended to be a proof of the method, for which one must refer to the original article.

Let X be a random variable with the cumulative distribution function $F(x)$, possessing a pdf $f(x)$, and a unique median which, without any loss of generality will be taken to be zero. Suppose that $m(z)$, the inverse function of $z(m) = 2F(m) - 1$, is for $-1 < z < 1$ uniquely defined and equal to a convergent series of powers of z; let

$$m(z) = a_1 z(1 + \sum_{p=1}^{\infty} c_p z^p). \tag{1}$$

Theorems 2 and 3 (pp. 597–599) of Chu and Hotelling imply that, if $c_p/p^k \rightarrow$ constant (which may be zero) for some integer k, then the series in (7) below are convergent for all integers n greater than some n_0. Further, under the same conditions, an argument that they used for the special case of the double exponential distribution (p. 599) can be generalized to show that the series in (7) are asymptotic in n.

Now, for the U-shaped, rectangular, double exponential, Cauchy and the exponential distributions these conditions are easily verified as general expressions for c_p are available. In fact, for the rectangular distribution the series terminates at the first term; for the double exponential and exponential $c_p = O(p^{-1})$; for the Cauchy $c_p = O(1)$; and for the U-shaped $c_p = O(a^p p^{-p-\frac{1}{2}})$, where a is a constant, so that $c_p/p^k \to 0$ for every k as $p \to \infty$. General expressions for c_p are not available for the normal and parabolic distributions. However, the normal distribution has been treated at length by Chu and Hotelling and shown to have the desired properties. Only in the case of the parabolic distribution we fail to verify the (sufficient) condition of their Theorem 3. The numerical calculations, however, indicate that the series (7) corresponding to this case are also convergent for $n \geqq 1$ and asymptotic in n. Let M denote the sample median of a sample of size $N = 2n + 1$. Then the pdf of M is

$$g(m) = [(2n+1)!/(n!)^2][F(m)]^n[1 - F(m)]^n f(m). \tag{2}$$

Therefore

$$EM^r = \frac{(2n+1)!}{(n!)^2} \int_{-\infty}^{\infty} m^r[F(m)]^n[1 - F(m)]^n f(m)\, dm. \tag{3}$$

Make the transformation

$$z(m) = 2F(m) - 1 = 2\int_0^m f(x)\, dx \tag{4}$$

in (3) and insert series (1) for $m(z)$, the inverse function of $z(m)$. Then

$$\begin{aligned} EM^r &= [B\,(n+1, \tfrac{1}{2})]^{-1} \int_{-1}^{1} m^r(z)(1 - z^2)^n\, dz \\ &= a_1^r\,[B\,(n+1, \tfrac{1}{2})]^{-1} \int_{-1}^{1} z^r(1 + \textstyle\sum c_p z^p)^r (1 - z^2)^n\, dz. \end{aligned} \tag{5}$$

We note that, for $r = 0, 1, 2, \cdots$,

$$\begin{aligned} &\int_{-1}^{1} z^{2r+1}(1 - z^2)^n\, dz = 0, \\ &\int_{-1}^{1} z^{2r}(1 - z^2)^n\, dz = \int_0^1 x^{r-\frac{1}{2}}(1 - x)^n\, dx = B\,(n+1, r+\tfrac{1}{2}). \end{aligned} \tag{6}$$

Expanding $(1 + \sum c_p z^p)^r$ in a multinomial expansion, integrating term by term with the help of (6), we obtain, for $r = 0, 1, 2, \cdots$,

$$\begin{aligned} EM^{2r} &= \frac{a_1^{2r} B(n+1, r+\frac{1}{2})}{B(n+1, \frac{1}{2})} \\ &\cdot \left[1 + \frac{2r+1}{2n+2r+3}\, b_2(r) + \frac{(2r+1)(2r+3)}{(2n+2r+3)(2n+2r+5)}\, b_4(r) + \cdots\right], \\ EM^{2r+1} &= \frac{(2r+1) a_1^{2r+1} B(n+1, r+\frac{3}{2})}{B(n+1, \frac{1}{2})} \\ &\cdot \left[c_1 + \frac{2r+3}{2n+2r+5}\, b_3(r) + \frac{(2r+3)(2r+5)}{(2n+2r+5)(2n+2r+7)}\, b_5(r) + \cdots\right]. \end{aligned} \tag{7}$$

Here, $b_2(r), b_3(r), \cdots$, can be determined in terms of $r, c_1, c_2, \cdots$. For example,

$$b_2(r) = r(2r - 1)c_1^2 + 2rc_2,$$

$$b_3(r) = \tfrac{1}{3}r(2r - 1)c_1^3 + 2rc_1c_2 + c_3.$$

There is little point in giving algebraic expressions for b's as it is more convenient to work them out numerically for each distribution.

Since, under the conditions stated earlier in this section, the series in (7) are asymptotic in n, if $S_k(n)$ is the sum of the first k terms and $R_k(n)$ is the remainder, we have

$$\lim_{n\to\infty} \frac{S_k(n) + R_k(n)}{S_k(n)} = 1,$$

which implies

$$\lim_{n\to\infty} \left| \frac{R_k(n)}{R_k(n) + S_k(n)} \right| = 0.$$

Thus the relative error tends to zero as $n \to \infty$ for every fixed k. We therefore expect the relative error to be broadly decreasing with increasing n. Moreover, if all the coefficient c's are non-negative, so are the b's, and for fixed r and k,

$$R_k(n + 1) \leqq R_k(n),$$

the equality holding if and only if the series terminates at the kth term, i.e., if and only if $R_k(n) = 0$. Thus, if $R_k(n) \neq 0$, the absolute error of approximation decreases monotonically for every fixed k.

In all the cases studied the coefficient c's are non-negative, excepting the U-shaped distribution. However, even for that distribution we find that, so far as our computations go, the absolute error is monotonically decreasing with increasing n.

We observe that, in some cases a unique expansion of $m(z)$ may readily be available for $|z| < 1$, e.g., for rectangular, Cauchy, U-shaped and exponential distributions; otherwise suppose that Maclaurin's series expansion of $z(m)$ exists. It is then given by

$$z(m) = \sum_{j=1}^{\infty} 2m^j \frac{f^{(j)}}{j!};$$

where $f^{(j)}$ is the jth derivative of $f(x)$ evaluated at zero. Inversion of this series may give the required expansion of $m(z)$. In some cases a unique expansion of $m(z)$ for $-1 < z < 1$ may not be possible, but different expansions may exist for different parts of the range of z, e.g., for the double exponential distribution. In such cases (7) will not hold and we will have to evaluate the integral (5) by splitting the range into as many parts as needed. We further observe that if $f(x)$ is symmetrical about the median the calculations are much simplified as $c_1 = c_3 = \cdots = 0$; also $EM^{2r+1} = 0$, whenever it exists, for $r = 0, 1, 2, \cdots$.

Note also the following. Let M_1 and M_2 be the medians of samples of the same size from two different symmetric populations F_1 and F_2. If $m_1(z)$ and $m_2(z)$ correspond to F_1 and F_2 respectively, and if

$$m_1(x) \leqq m_2(z)$$

for $0 < z < 1$, where strict inequality holds over an interval of non-zero length, then

$$EM_1^{2r} < EM_2^{2r}, \qquad r = 1, 2, \cdots .$$

Compare, for example, the Cauchy and the U-shaped distributions. For the U-shaped distribution (F_1), $m_1(z) = \sin(\frac{1}{2}\pi z)$, and for the Cauchy distribution (F_2), $m_2(z) = \tan(\frac{1}{2}\pi z)$, and hence $EM_1^{2r} < EM_2^{2r}$, $r = 1, 2, \cdots$.

3. Notation. For brevity we shall use the following notation:

$$A_k(x) = [x(x+2)(x+4) \cdots (x+2k-2)]^{-1}.$$

Thus, for example, $A_1(N+2) = (N+2)^{-1}$, $A_2(N+4) = [(N+4)(N+6)]^{-1}$,

4. Applications.

4.1. *U-shaped (incomplete beta) distribution.* Let

$$f(x) = \pi^{-1}(1 - x^2)^{-\frac{1}{2}},$$

if $x^2 < 1$; 0, otherwise. We have

$$m = \sin(\tfrac{1}{2}\pi z) = (\tfrac{1}{2}\pi z)[1 - (3!)^{-1}(\tfrac{1}{2}\pi z)^2 + (5!)^{-1}(\tfrac{1}{2}\pi z)^4 - \cdots];$$

$$EM^{2r+1} = 0,$$

$$\begin{aligned} EM^{2r} = \left(\frac{\pi^2}{4}\right)^r \frac{B(n+1, r+\frac{1}{2})}{B(n+1, \frac{1}{2})} \Bigg[1 &- \frac{r(2r+1)\pi^2}{12} A_1(N+2r+2) \\ &+ \left(\frac{r}{960} + \frac{r(2r+1)}{576}\right)(2r+1)(2r+3)\pi^4 A_2(N+2r+2) \\ &- \left(\frac{r}{32 \times 7!} + \frac{r(2r-1)}{192 \times 5!} + \frac{(r-1)r(2r-1)}{20736}\right) \\ &\cdot (2r+1)(2r+3)(2r+5)\pi^6 A_3(N+2r+2) + \cdots \Bigg]. \end{aligned}$$

In particular

$$\begin{aligned} EM^2 \cong\, & 2.4674011\, A_1(N+2)[1 - 2.4674011\, A_1(N+4) \\ & + 4.0587121\, A_2(N+4) - 5.0072354\, A_3(N+4) + 4.9419432\, A_4(N+4)]. \end{aligned} \tag{9}$$

This is the only case in which the series for m is of alternating nature; in all other cases, as we shall see, all the non-zero coefficients of the powers of z are positive.

The referee has pointed out that EM^2 is expressible in terms of the Bessel function $J_{n+\frac{1}{2}}(x)$ evaluated at $x = \pi$. Thus, from (5),

$$EM^2 = [B(n+1, \tfrac{1}{2})]^{-1} \int_{-1}^{1} \sin^2(\pi z/2)(1-z^2)^n \, dz$$

$$(10) \qquad = \tfrac{1}{2} - \Gamma(n+\tfrac{3}{2})\,[2\pi^{\frac{1}{2}} n!]^{-1} \int_{-1}^{1} \cos \pi z (1-z^2)^n \, dz$$

$$= \tfrac{1}{2} - 2^{n-\frac{1}{2}} \pi^{-n-\frac{1}{2}} \Gamma(n+\tfrac{3}{2})\, J_{n+\frac{1}{2}}(\pi).$$

In Table 1 the approximate var M as calculated from (9) is compared with the exact var M calculated from (10). var M/var $\bar{X}$ is also tabulated, where $\bar{X}$ is the sample mean.

TABLE 1
U-shaped distribution, var M, *and* var M/var $\bar{X}$
(var X = 0.5)

N	1	3	5	7	11	17	31	∞
Approx. var M	0.500071	0.348033	0.269020	0.219838	0.161413	0.115644	0.069726	0
Exact var M	0.500000	0.348018	0.269015	0.219837				0
Relative error	0.00014	0.00004	0.00002	4.5×10^{-6}				0
var M/var $\bar{X}$	1	2.0881	2.6902	3.0777	3.5511	3.9319	4.3230	4.9348

4.2. *Rectangular distribution.* Let $f(x) = 1, -\frac{1}{2} \leqq x \leqq \frac{1}{2}$; 0, otherwise. It is easily seen that $m = \frac{1}{2}z$, so that for $r = 0, 1, 2, \cdots$

$$(11) \qquad EM^{2r+1} = 0, \qquad EM^{2r} = 4^{-r} B(n+1, r+\tfrac{1}{2})/B(n+1, \tfrac{1}{2})$$

These expressions are exact. In particular $EM^2 = \frac{1}{4}(N+2)^{-1}$. In Table 2 var M and var M/var $\bar{X}$ are tabulated for selected values of N.

TABLE 2
Rectangular distribution. Exact var M *and* var M/var $\bar{X}$
(var X = 0.08333)

N	1	3	5	7	11	17	31	∞
var M	0.08333	0.05000	0.03571	0.02778	0.01923	0.01316	0.00758	0
var M/var $\bar{X}$	1.0000	1.8000	2.1428	2.3333	2.5385	2.6842	2.8182	3

4.3. *Parabolic distribution.* Let $f(x) = (\frac{3}{4})(1-x^2)$, for $x^2 \leqq 1$; 0, otherwise. We obtain

$$(12) \qquad z = 3m/2 - m^3/2,$$

and therefore

$$m = (2z/3)[1 + 1/3(2z/3)^2 + 1/3(2z/3)^4 + 4/9(2z/3)^6 + \cdots].$$

It can easily be seen, replacing m by the series (4) in (12) that $c_{2k-1} = 0$, and c_{2k} is a polynomial in $a_1, c_2, c_4, \cdots, c_{2k-2}$, with all the coefficients positive. Since a_1 and c_2 are positive, all c_{2k} are positive.

We have $EM^{2r+1} = 0$,

$$EM^{2r} = \left(\frac{4}{9}\right)^r \frac{B(n+1, r+\frac{1}{2})}{B(n+1, \frac{1}{2})} \left[1 + \frac{8}{27} r(2r+1) A_1(N+2r+2)\right.$$

$$+ \left(\frac{32r}{243} + \frac{16r(2r-1)}{729}\right)(2r+1)(2r+3)A_2(N+2r+2)$$

$$\left. + \frac{128r(2r^2+3r+10)}{19683}(2r+1)(2r+3)(2r+5)A_3(N+2r+2) + \cdots\right].$$

In particular

$$EM^2 \cong 4/9\, A_1(N+2)[1 + 8/9\, A_1(N+4) + 2.3045267\, A_2(N+4) + 10.242341\, A_3(N+4) + 65.095768\, A_4(N+4)]. \tag{13}$$

In Table 3 var M as calculated from (13) is compared with the exact value for selected values of N. The latter values are taken from Rider [4].

TABLE 3

Parabolic distribution, var M *and* var M/var $\bar{X}$

(var $X = 0.2$)

N	1	3	5	7	11	17	31	∞
Approx. var M	0.1918	0.1054	0.0720	0.0545	0.0366	0.0245	0.0138	0
Exact var M	0.2000	0.1065	0.0722	0.0546				0
Relative error	0.036	0.010	0.003	0.002				0
var M/var $\bar{X}$	1	1.597	1.805	1.911	2.035	2.083	2.145	2.222

4.4. *Normal distribution*. Let

$$f(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2}, \qquad -\infty \leqq x \leqq \infty.$$

Writing $f = f(0)$, and $f^{(j)} = f^{(j)}(0)$, we have $f^{(2j+1)} = 0, j = 0, 1, 2, \cdots$,

$$f = (2\pi)^{-\frac{1}{2}}, \quad f^{(2)} = -f, \quad f^{(4)} = 3f, \quad f^{(6)} = -15f, \quad f^{(8)} = 105f.$$

Hence

$$\left(\frac{\pi}{2}\right)^{\frac{1}{2}} z = m - \frac{m^3}{3!} + \frac{3m^5}{5!} - \frac{15m^7}{7!} + \frac{105m^9}{9!} + \cdots.$$

The inversion of this series gives

$$m = \left(\frac{\pi}{2}\right)^{\frac{1}{2}} z\left[1 + \frac{\pi}{12}z^2 + \frac{7\pi^2}{480}z^4 + \frac{127\pi^3}{8!}z^6 + \frac{4369\pi^4}{16 \times 9!}z^8 + \cdots\right].$$

The referee has pointed out that from equation (4.11), Table 1, and the recurrence relation below Table 1 of J. G. Saw's article [7, p. 215], it is obvious that all the non-zero coefficients of the powers of z are positive.

We find

$$EM^{2r+1} = 0$$

$$EM^{2r} = \left(\frac{\pi}{2}\right)^r \frac{B(n+1, r+\frac{1}{2})}{B(n+1, \frac{1}{2})}\Bigg[1 + \frac{r(2r+1)\pi}{6} A_1(N+2r+2) + \left(\frac{7r}{240} + \frac{r(2r-1)}{144}\right)(2r+1)(2r+3)\pi^2 A_2(N+2r+2) + \left(\frac{254r}{8!} + \frac{7r(2r-1)}{2880} + \frac{2(r-1)r(2r-1)}{5184}\right) \cdot (2r+1)(2r+3)(2r+5)\pi^3 A_3(N+2r+2) + \cdots\Bigg].$$

In particular

$$\begin{aligned} EM^2 &\cong 1.5707963\ A_1(N+2)[1 + 1.5707963\ A_1(N+4) \\ &+ 5.3460357\ A_2(N+4) + 28.422420\ A_3(N+4) \\ &+ 206.43625\ A_4(N+4)], \end{aligned} \tag{14}$$

$$\begin{aligned} EM^4 &\cong 7.4022033\ A_2(N+2)[1 + 5.2358977\ A_1(N+6) \\ &+ 34.543615\ A_2(N+6) + 288.09999\ A_3(N+6) \\ &+ 2936.2077\ A_4(N+6)]. \end{aligned} \tag{15}$$

In Table 4 var M and EM^4 as calculated from (14) and (15) are compared with the exact values for selected values of N. The exact values of var M are taken from Teichroew [8], and of EM^4 from Hojo [3]. The exact values of var M/var $\bar{X}$ are also tabulated.

TABLE 4

Normal distribution, var M, EM^4, *and* var M/var $\bar{X}$

(var X = 1, EX^4 = 3)

N	1	3	5	7	11	17	31	∞
Approv. var M	0.84649	0.43140	0.28304	0.20928	0.13697	0.09002	0.04996	0
Exact var M	1.00000	0.44867	0.28683	0.21045	0.13716	0.09005		0
Relative error	0.1535	0.0385	0.0132	0.0056	0.0014	0.0003		0
Approx. EM^4			0.2239	0.1281				0
Exact EM^4			0.2495	0.1341				0
Relative error			0.103	0.045				0
var M/var $\bar{X}$	1	1.3460	1.4342	1.4731	1.5088	1.5308	1.5488	1.5708

4.5. *Double exponential (Laplace) distribution.* Let $f(x) = \frac{1}{2}e^{-|x|}$, $-\infty \leqq$

$x \leqq \infty$. We find

$$z(m) = \begin{cases} 1 - e^{-m}, & \text{if } m \geqq 0; \\ e^{m} - 1, & \text{if } m \leqq 0. \end{cases}$$

Obviously $EM^{2r+1} = 0$, $r = 0, 1, 2, \cdots$, and

$$EM^{2r} = 2[B(n+1, \tfrac{1}{2})]^{-1} \int_0^1 [-\ln(1-z)]^{2r}(1-z^2)^n \, dz$$

$$= [B(n+1, \tfrac{1}{2})]^{-1} \int_0^1 \left(\sum_{j=1}^{\infty} j^{-1} x^{\frac{1}{2}j}\right)^{2r} (1-x)^n x^{-\frac{1}{2}} \, dx$$

$$= \frac{B(n+1, r+\frac{1}{2})}{B(n+1, \frac{1}{2})} \left[1 + \frac{rB(n+1, r+1)}{B(n+1, r+\frac{1}{2})} + \left(\frac{2r}{3} + \frac{r(2r-1)}{4}\right) \frac{B(n+1, r+\frac{3}{2})}{B(n+1, r+\frac{1}{2})} + \cdots\right].$$

In particular, writing

$$C_n = 2^{-2n-2}(2n+3)![(n+2)!(n+1)!]^{-1},$$

We have

$$\begin{aligned} EM^2 \cong A_1(2n+3)[1 + C_n + 2.75\, A_1(2n+5) + 15/9\, C_n A_1(n+3) \\ +11.416667\, A_2(2n+5) + 4.2\, C_n A_2(n+3) + 68.0625\, A_3(2n+5) \\ +14.495238\, C_n A_3(n+3) + 534.675\, A_4(2n+5)]. \end{aligned} \tag{16}$$

Since $C_n \sim 2\pi^{-\frac{1}{2}} n^{-\frac{1}{2}}$, the series in the square brackets proceeds asymptotically with terms of order $1, n^{-\frac{1}{2}}, n^{-1}, n^{-\frac{3}{2}}, \cdots$; hence the necessity of taking nine terms to reach the term of relative order n^{-4}. For $N = 1, 3$, and 5, i.e., for $n = 0$, 1, and 2, the exact values of var M are available from Sarhan [6]. In Table 5 var M as calculated from (16) is tabulated for selected values of N and the initial three values are compared with the exact values. var M/var $\bar{X}$ is also given.

TABLE 5

Double exponential distribution. var M *and* var M/var $\bar{X}$

(var $X = 2$)

N	1	3	5	7	9	17	∞
Approx. var M	1.28564	0.56476	0.33499	0.23062	0.17314	0.08325	0
Exact var M	2	0.63889	0.35118				0
Relative error	0.3572	0.1160	0.0461				0
var M/var $\bar{X}$	1	0.95833	0.87795	$\cong$0.81	$\cong$0.78	$\cong$0.71	0.5

4.6. *Cauchy distribution.* Let $f(x) = \pi^{-1}(1 + x^2)^{-1}$, $-\infty \leqq x \leqq \infty$. It is easily found that

$$m = \tan\frac{\pi z}{2} = \frac{\pi z}{2}\left[1 + \frac{1}{3}\left(\frac{\pi z}{2}\right)^2 + \frac{2}{15}\left(\frac{\pi z}{2}\right)^4 + \frac{17}{315}\left(\frac{\pi z}{2}\right)^6 + \frac{62}{2835}\left(\frac{\pi z}{2}\right)^8 + \cdots\right].$$

From the known expansion of tan z, we know that all the non-zero coefficients of the powers of z are positive. If $n \geqq 2r + 1$, $EM^{2r+1} = 0$; and when $n \geqq 2r$,

$$\begin{aligned} EM^{2r} \cong \left(\frac{\pi^2}{4}\right)^r \frac{B(n+1, r+\frac{1}{2})}{B(n+1,\frac{1}{2})} \Bigg[1 &+ \frac{r(2r+1)\pi^2}{6} A_1(N+2r+2) \\ &+ \left(\frac{r}{60} + \frac{r(2r-1)}{144}\right)(2r+1)(2r+3)\pi^4 A_2(N+2r+2) \\ &+ \left(\frac{17r}{10080} + \frac{r(2r-1)}{720} + \frac{(r-1)r(2r-1)}{2592}\right) \\ &\qquad \cdot (2r+1)(2r+3)(2r+5)\pi^6 A_3(N+2r+2) + \cdots \Bigg]. \end{aligned}$$

In particular, for $N \geqq 5$,

$$\begin{aligned} EM^2 \cong 2.4674011\, A_1(N+2)[1 &+ 4.9348022\, A_1(N+4) \\ &+ 34.499053\, A_2(N+4) + 310.44860\, A_3(N+4) \\ &+ 3414.8826\, A_4(N+4)]. \end{aligned} \tag{17}$$

The coefficient of the next term in the square brackets is 44393.397.

In Table 6 the approximate value of var M as calculated from (17) is compared with the exact value for selected values of N. The latter is given by Rider [5] for $N = 5(2)31$.

TABLE 6

Cauchy distribution, var M

N	5	7	11	17	31	∞
Approx. var M	0.82148	0.52864	0.29645	0.17435	0.08789	0
Exact var M	1.22125	0.61208	0.30680	0.17562	0.08794	0
Relative error	0.3273	0.1363	0.0337	0.0072	0.0006	0

4.7. *Exponential distribution.* We take the exponential distribution in the form $f(x) = \frac{1}{2}e^{-x}$, for $-\ln 2 \leqq x \leqq \infty$; 0, otherwise, so that the population median is zero. It is easily found that

$$m = -\ln(1 - z) = z(1 + z/2 + z^2/3 + z^3/4 + \cdots),$$

and all the coefficients of the powers of z are positive;

$$EM^{2r+1} = \frac{(2r+1)B(n+1, r+\frac{3}{2})}{2B(n+1,\frac{1}{2})}$$

$$\cdot\left[1+\left(\frac{1}{2}+\frac{2r}{3}+\frac{r(2r-1)}{12}\right)(2r+3)A_1(N+2r+4)+\cdots\right],$$

$$EM^{2r}=\frac{B(n+1,r+\frac{1}{2})}{B(n+1,\frac{1}{2})}\left[1+\left(\frac{2r}{3}+\frac{r(2r-1)}{4}\right)(2r+1)A_1(N+2r+2)\right.$$

$$+\left(\frac{2r}{5}+\frac{11r(2r-1)}{36}+\frac{r(r-1)(2r-1)}{6}\right.$$

$$\left.\left.+\frac{r(r-1)(2r-1)(2r-3)}{96}\right)(2r+1)(2r+3)A_2(N+2r+2)+\cdots\right].$$

In particular

$$EM \cong 1/2\, A_1(N+2)[1+1.5\, A_1(N+4)+5\, A_2(N+4)$$
$$+\, 26.25\, A_3(N+4)+189\, A_4(N+4)],$$

$$\text{(18)}\quad EM^2 \cong A_1(N+2)[1+2.75\, A_1(N+4)+11.416667\, A_2(N+4)$$
$$+\, 68.0625\, A_3(N+4)+534.675\, A_4(N+4)].$$

The values of EM and var $M = EM^2 - (EM)^2$ as calculated from (18) are compared with their exact values for selected values of N in Table 7. The latter are taken from Gupta [2].

TABLE 7

Exponential distribution. EM, var M, *and* var M/var $\bar{X}$

($EX = 1 - \ln 2$, var $X = 1$)

N	1	3	5	7	9	∞
Approx. EM	0.2635	0.1353	0.0891	0.0660	0.0524	0
Exact EM	0.3069	0.1402	0.0902	0.0664	0.0525	0
Relative error	0.141	0.035	0.012	0.006	0.002	0
Approx. var M	0.6795	0.3275	0.2066	0.1486	0.1154	0
Exact var M	1.0000	0.3611	0.2136	0.1507	0.1162	0
Relative error	0.321	0.093	0.033	0.014	0.007	0
var M/var $\bar{X}$	1	1.083	1.068	1.055	1.045	1

5. Acknowledgment. Thanks are due to the referee for many valuable suggestions, especially for providing the exact expression for the variance of the sample median for U-shaped population, and for drawing the author's attention to J. G. Saw's paper.

REFERENCES

[1] JOHN T. CHU AND HAROLD HOTELLING, "The moments of the sample median," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 593–606.

[2] SHANTI S. GUPTA, "Order statistics from the gamma distribution," *Technometrics*, Vol. 2 (1960), pp. 243–262.

[3] TOKISHIGE HOJO, "Distribution of the median, quartiles and interquartile distance in samples from a normal population," *Biometrika*, Vol. 23 (1931), pp. 315–360.
[4] PAUL R. RIDER, "Variance of the median of small samples from several special populations," *J. Amer. Stat. Assn.*, Vol. 55 (1960), pp. 148–150.
[5] PAUL R. RIDER, "Variance of the median of samples from a Cauchy distribution," *J. Amer. Stat. Assn.*, Vol. 55 (1960), pp. 322–323.
[6] A. E. SARHAN, "Estimation of the mean and standard deviation by order statistics," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 317–328.
[7] J. G. SAW, "Moments of sample moments of censored samples from a normal population," *Biometrika*, Vol. 45 (1958), pp. 211–221.
[8] D. TEICHROEW, "Tables of expected values of order statistics and products of order statistics for samples of size twenty and less from the normal distribution," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 410–426.