M A T H E M A T I C A L   C E N T R E

2e Boerhaavestraat 49

AMSTERDAM

S t a t i s t i c a l   D e p a r t m e n t

Direction: Prof.Dr D. van Dantzig

Chief of Statistical Consultation: Prof.Dr J. Hemelrijk

Report SP 48

Simple approximate formulas for the median of order statistics

by

A. Benard and Emily C. Bos-Levenbach

October
1955

Simple approximate formulas for the median of order statistics.[1)]

by
A. Benard and Emily C. Bos-Levenbach

Introduction

Let $\underline{x}_1, \ldots, \underline{x}_n$ [2)] be the n order statistics of a random sample of size n from a population with a continuous probability distribution. Let $F(x)$ denote the cumulative distribution function of the sampled variable and let

(1) $$\underline{y}_i = F(\underline{x}_i) \qquad (i = 1, \ldots, n).$$

The following approximate formula will be derived for the median y_i^* of $\underline{y}_i$:

(2) $$y_i^* = \mathrm{Med}\, \underline{y}_i \approx \frac{i - 0.3}{n + 0.4}$$

This formula holds for small values of i or $n-i$. For intermediate values the formula

(3) $$y_i^* \approx \frac{i - 1/3}{n + 1/3}$$

is more accurate.

The median x_i^* of the order statistic $\underline{x}_i$ is approximately equal to

(4) $$x_i^* = \mathrm{Med}\, \underline{x}_i \approx F^{-1}\left(\frac{i - 0.3}{n + 0.4}\right)$$

or

(5) $$x_i^* \approx F^{-1}\left(\frac{i - 1/3}{n + 1/3}\right)$$

respectively, where F^{-1} denotes the inverse function of F.

Proof.

The cumulative distributionfunction $G_i(y)$ of $\underline{y}_i$ $(i = 1, \ldots, n)$ is given by

(6) $$G_i(y) = P[\underline{y}_i \leq y] = \sum_{k=i}^{n} \binom{n}{k} y^k (1-y)^{n-k} = \frac{n!}{(i-1)!(n-i)!} \int_0^y u^{i-1} (1-u)^{n-i}\, du \quad (0 \leq y \leq 1).$$

Thus for given values of n and i the median y_i^* of $\underline{y}_i$ can be found by means of tables of the incomplete bêta function (e.g. C.H. THOMPSON [3]).

---

1) Report SP 48 of the Statistical Department of the Mathematical Centre, Amsterdam. Head of the Department: Prof.Dr D. VAN DANTZIG.

2) The random character of a variable is denoted by underlining its symbol. The same symbol, not underlined, can then be used for values which may be assumed by this random variable.

To derive a simple approximate formula we first remark that the following relation holds:

(7) $\mathcal{G}_i(y) = 1 - \mathcal{G}_{n+1-i}(1-y).$

This follows from (6):

$$\mathcal{G}_i(y) = \sum_{k=i}^{n} \binom{n}{k} y^k (1-y)^{n-k} = 1 - \sum_{k=0}^{i-1} \binom{n}{k} y^k (1-y)^{n-k} =$$

$$= 1 - \sum_{k=n+1-i}^{n} \binom{n}{k} (1-y)^k y^{n-k} = 1 - \mathcal{G}_{n+1-i}(1-y).$$

For the medians y_i^* and y_{n+1-i}^* we have

(8) $\mathcal{G}_i(y_i^*) = \mathcal{G}_{n+1-i}(y_{n+1-i}^*) = 1/2$

and according to (7)

$$\mathcal{G}_i(y_i^*) = 1 - \mathcal{G}_{n+1-i}(1-y_i^*) = 1/2.$$

Thus

$$\mathcal{G}_{n+1-i}(1-y_i^*) = \mathcal{G}_{n+1-i}(y_{n+1-i}^*) = 1/2$$

and

(9) $y_{n+1-i}^* = 1 - y_i^*.$

We may now write

(10) $y_i^* = \dfrac{i-a}{n+b},$

where a and b are functions of n and i. Substitution of (10) into (9) gives

$$\frac{n+1-i-a}{n+b} = 1 - \frac{i-a}{n+b}$$

or

(11) $b = 1 - 2a,$

and (10) may be written in the following form

(12) $y_i^* = \dfrac{i-a}{n+1-2a}$

a still being a function of n and i.

Substituting (12) into (8) gives

(13) $\sum_{k=i}^{n} \binom{n}{k} \left(\dfrac{i-a}{n+1-2a}\right)^k \left(1 - \dfrac{i-a}{n+1-2a}\right)^{n-k} = 1/2$

and the values of a satisfying this relation may be found by means of a table of the incomplete bêta function. Table I gives these values of a for some small values of n and i.

Table I

$a.10^3$ for small values of n and i.

i \ n	2	3	4	5	6	7	8
1	293	298	300	301	302	303	303
2			312	305	316	317	318
3					319	320	321
4							322

For small n and for $i=1$ the value $a = 0.3$ seems to be a fairly good approximation.

As a next step the asymptotic behaviour of a for $n \to \infty$ will be investigated. The relation (13) may also be written as follows:

$$\tfrac{1}{2} = \sum_{k=0}^{i-1} \binom{n}{k} \left(\frac{i-a}{n+1-2a}\right)^k \left(1 - \frac{i-a}{n+1-2a}\right)^{n-k}.$$

For $n \to \infty$ and i constant

$$\lim \left(1 - \frac{i-a}{n+1-2a}\right)^{n-k} = e^{-(i-a)}$$

and

$$\lim \binom{n}{k} (n+1-2a)^{-k} = 1/k!$$

thus for $n \to \infty$ we may write

$$(14) \quad e^{-(i-a)} \sum_{k=0}^{i-1} \frac{(i-a)^k}{k!} = \tfrac{1}{2}.$$

In order to find a for given i and $n \to \infty$ we must therefore find a Poisson distribution with mean $i-a$ and median i. Using the tables of E.C. MOLINA [2] table II was computed.

Table II

a for different values of i and $n \to \infty$.

i	a
1	0.307
2	0.321
5	0.329
10	0.331
50	0.332
100	0.333

For $i = 1$ the limiting value of α for $n \to \infty$ follows from (14):

$$e^{-(1-\alpha)} = 1/2 ,$$

giving

(15) $\alpha = 1 - \ln 2$.

For $n \to \infty$ and $i/n =$ constant another limiting value of α may be derived. This may be done by substituting

$$i/n = \alpha + \varepsilon_n/n$$

into (12) and using an approximation for the binomial distribution given by J.V. USPENSKY [4] p. 129. The result is $\alpha = 1/3$.[3]

Finally, to check the degree of approximation, for $i = 1$ and a few small values of n, the exact value of $G_i\{(i-a)/(n+1-2a)\} = G_1\{0.7/(n+0.4)\}$ was computed.

The results, given in table III, are favourable for the value of α chosen.

Table III

Exact values of $G_1\{0.7/(n+0.4)\}$.

n	G
2	0.4983
3	0.4992
4	0.5000
5	0.5005

The approximation is satisfying and has, moreover, been checked for $n = 10$ and 15 by computing the exact values of the median for $i = 1, \ldots, 10$ and $1, \ldots, 15$ respectively and computing these with the corresponding values of $(i-0.3)/(n+0.4)$.

The difference proved to be smaller than 1% for every value investigated.

Remark

Formula (2) may be put into use in the plotting of observations on probability paper. It enables one to plot the points in such

---

3) This computation, suggested by Prof.Dr D. VAN DANTZIG, was executed by H. KESTEN and TH.J, RUNNENBURG, assistants of the Statistical Department of the Mathematical Centre, Amsterdam.

a way that every point plotted has approximately equal chance of falling above or below the true line. Cf. A. BENARD and E.C. BOS-LEVENBACH [1] .

Acknowledgement

The authors want to thank Prof.Dr J. Hemelrijk for many helpful suggestions.

Literature

[1] A. Benard and E.C. Bos-Levenbach, The plotting of observations on probability paper, Report SP 30A of the Statistical Department of the Mathematical Centre, Amsterdam. Statistica, 7 (1953), 3, p. 163-173.

[2] E.C. Molina, Poisson's exponential binomial limit, D. van Nostrand, Comp., Inc., N.Y.(1954).

[3] C.M. Thompson, Tables of percentage points of the incomplete bèta-function, Biometrika, 32 (1941) p. 168-181.

[4] J.V. Uspensky, Introduction to mathematical probability (6th edition 1937). Mc Graw Hill Book Cy, New York and London.
