stichting
mathematisch
centrum

AFDELING MATHEMATISCHE STATISTIEK SW 59/78 JULI
(DEPARTMENT OF MATHEMATICAL STATISTICS)

R. KAAS & J.M. BUHRMAN

A NOTE ON THE MEDIAN OF THE BINOMIAL DISTRIBUTION

Preprint

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O).

AMS(MOS) subject classification scheme (1970): 62E01

A note on the median of the binomial distribution*)

by

R. Kaas & J.M. BUHRMAN

SUMMARY

The location of the median of the binomial distribution is studied. An integer k is shown to be a median of the binomial-(n,p) distribution if it does not differ from np by more than min{p,1-p}. Neither the inequality "mean < median < mode" nor its reverse holds if median and mode differ (see [4]).

KEY WORDS & PHRASES: *Binomial distribution, median, mode.*

*) This report will be submitted for publication elsewhere.

1. INTRODUCTION

In two papers ([2], [3]) Neumann gives different proofs for the fact that the median of a binomial-(n,p) distribution equals np if np is an integer. We shall show that the median equals an integer k, if $|k - np| \leq \min\{p, 1-p\}$. First some definitions are given. A weak median of a random variable X will be a number x satisfying $P(X \geq x) \geq \frac{1}{2}$ and $P(X \leq x) \geq \frac{1}{2}$. A strong median is a number x for which $P(X \geq x - d) > \frac{1}{2}$ and $P(X \leq x + d) > \frac{1}{2}$ for all $d > 0$. Clearly, a strong median is unique. If a random variable has no strong median, its set of weak medians is a closed interval. Let, for k integer,

$$(1) \qquad b(k;n,p) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$(2) \qquad B(k;n,p) = \sum_{i=0}^{k} b(i;n,p)$$

UHLMANN [5] gives the following inequalities, cited in [1].

$$(3) \qquad B(k;n,\tfrac{k}{n-1}) > \tfrac{1}{2} > B(k;n,\tfrac{k+1}{n+1}) \qquad \text{for } k < \tfrac{1}{2}(n-1)$$

$$(4) \qquad B(k;n,\tfrac{k+1}{n+1}) > \tfrac{1}{2} > B(k;n,\tfrac{k}{n-1}) \qquad \text{for } k > \tfrac{1}{2}(n-1)$$

Obviously

$$(5) \qquad B(\tfrac{1}{2}(n-1);n,\tfrac{1}{2}) = \tfrac{1}{2} \qquad \text{if } n \text{ is odd}$$

2. THE MEDIAN

THEOREM 1. *If* X *has a binomial-*(n,p) *distribution and if the integer* k *satisfies*

$$(6) \qquad |k - np| \leq \min\{p, 1-p\}$$

then k *is a median of the distribution of* X*. The value* k *is a strong median unless* $p = \frac{1}{2}$.

PROOF. First the case $p = \frac{1}{2}$ is considered. If n is even, the only integer k satisfying (6) is $\frac{1}{2}$ n, which is obviously the strong median. If n is odd, both $k = \frac{1}{2}(n-1)$ and $k = \frac{1}{2}(n+1)$ satisfy (6) and the set of weak medians is the interval in between. Now let $p < \frac{1}{2}$. We shall show $P(X \geq k) > \frac{1}{2}$ and $P(X \leq k) > \frac{1}{2}$, which implies that k is the strong median. Note that $B(h;n,p)$ is decreasing in p. For $p < \frac{1}{2}$ condition (6) becomes

$$(7a) \qquad (n-1)p \leq k \leq (n+1)p$$

or, equivalently,

$$(7b) \qquad \frac{k}{n+1} \leq p \leq \frac{k}{n-1}$$

Now

$$P(X \geq k) = 1 - B(k-1;n,p) \geq 1 - B(k-1;n,\tfrac{k}{n+1}) > \frac{1}{2}$$

by the right hand inequality of (3), since $k-1 < \frac{1}{2}(n-1)$, because $k \leq (n+1)p < \frac{1}{2}(n+1)$.

$$P(X \leq k) = B(k;n,p) \geq B(k;n,\tfrac{k}{n-1}) > \frac{1}{2} \qquad \text{if } k < \frac{1}{2}(n-1)$$

Moreover, for $p < \frac{1}{2}$ we have $P(X \leq k) > \frac{1}{2}$ if either $k = \frac{1}{2}$ n and n is even, or $k = \frac{1}{2}(n-1)$ and n is odd. Since (7a) implies $k \leq (n+1)p < \frac{1}{2}(n+1)$, this completes the proof for the case $p < \frac{1}{2}$.

Next $p > \frac{1}{2}$ is considered. Let Y have a binomial-(n,q) distribution (with $q = 1-p$), and let $k' = n-k$. Then (6) becomes

$$(8) \qquad |k' - nq| \leq q$$

The preceding part of the proof shows that k' is the strong median of the distribution of Y, so k is the strong median of the distribution of X.

REMARK. By taking the sum of the lengths of the p-intervals in which the strong median is found by (7b), it can be shown that the theorem gives the median in about half the cases.

3. THE RELATIVE LOCATION OF THE MEAN, MEDIAN AND MODE

Let, for fixed n, the numbers $0 = p_0 < p_1 < \dots < p_n < p_{n+1} = 1$ be defined by

$$B(k-1;n,p_k) = \frac{1}{2} \qquad k = 1,2,\dots,n$$

This definition of p_k makes, that the strong median m_n is the following step-function of p.

$$m_n(p) = k \qquad \text{if } p \in (p_k, p_{k+1}) \text{ (see fig. 1)} \tag{9}$$

$m_n(p_k)$ is undefined, the binomial-(n,p_k) distribution has the interval $[k-1,k]$ as a set of weak medians. By considering (7b) it can be seen that theorem 1 is equivalent to statement (10) if n is odd, and to statement (11) if n is even.

$$p_1 < \frac{1}{n+1} < \frac{1}{n-1} < p_2 < \frac{2}{n+1} < \frac{2}{n-1} < p_3 < \dots \tag{10}$$

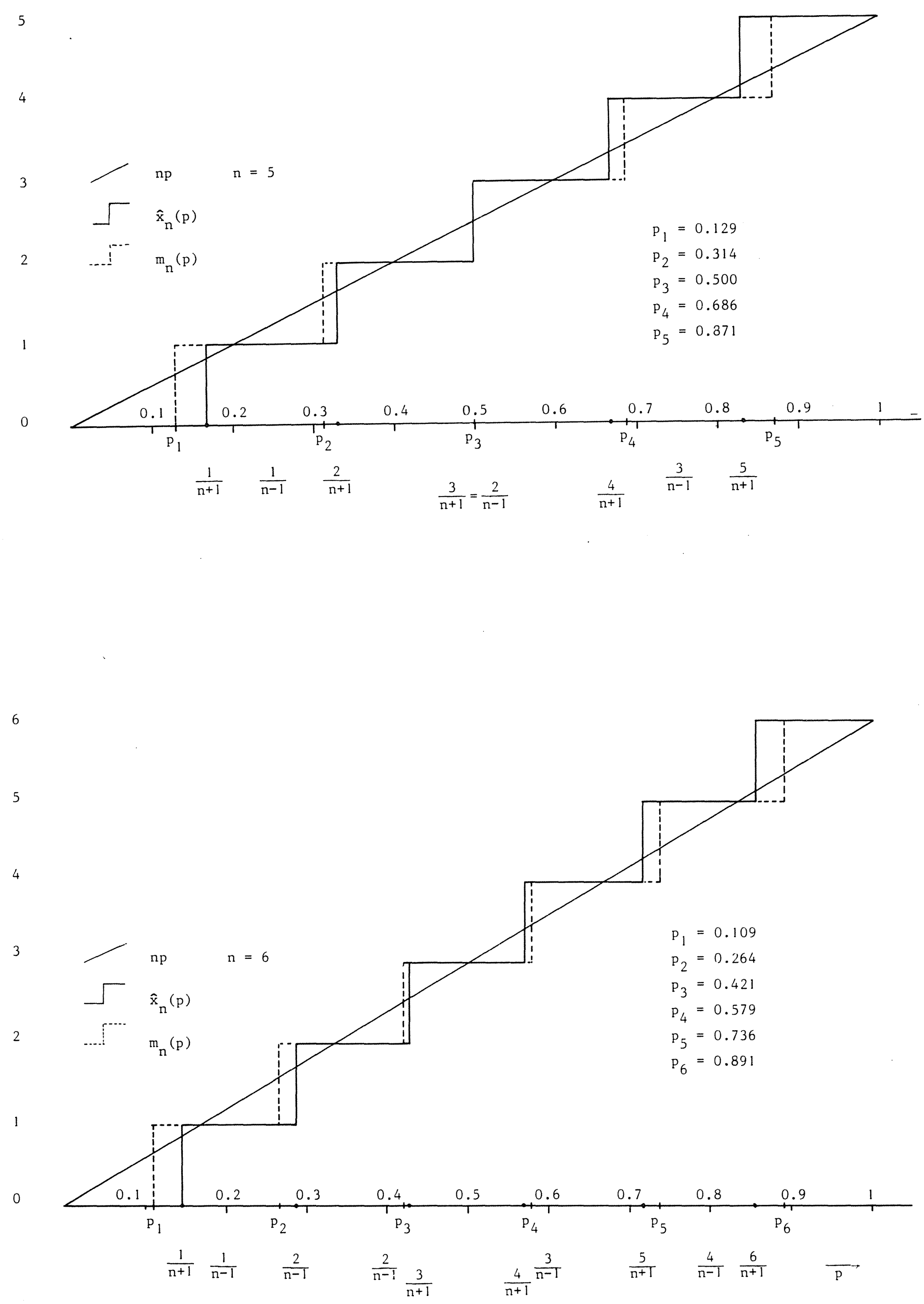
$$\dots < p_{\frac{n-1}{2}} < \frac{n-1}{2(n+1)} < \frac{1}{2} = p_{\frac{n+1}{2}} = \frac{1}{2} < \frac{n+3}{2(n+1)} < p_{\frac{n+3}{2}} < \dots$$

$$\dots < p_{n-2} < \frac{n-3}{n-1} < \frac{n-1}{n+1} < p_{n-1} < \frac{n-2}{n-1} < \frac{n}{n+1} < p_n$$

$$p_1 < \frac{1}{n+1} < \frac{1}{n-1} < p_2 < \frac{2}{n+1} < \frac{2}{n-1} < p_3 < \dots \tag{11}$$

$$\dots < p_{\frac{n}{2}} < \frac{n}{2(n+1)} < \frac{n+2}{2(n+1)} < p_{\frac{n}{2}+1} < \dots$$

$$\dots < p_{n-2} < \frac{n-3}{n-1} < \frac{n-1}{n+1} < p_{n-1} < \frac{n-2}{n-1} < \frac{n}{n+1} < p_n$$



<u>Figure 1</u> Mean, median and mode as functions of p for n = 5 and for n = 6.

Notice, that

(12) $$p_n = (\tfrac{1}{2})^{\frac{1}{n}}$$

(13) $$p_i + p_{n+1-i} = 1 \qquad \text{for all } i = 1,2,\ldots,n$$

The mode $\hat{x}_n$ of a binomial-(n,p) distribution equals $[(n+1)p]$. Therefore the mode is given by the following step-function in p.

(14) $$\hat{x}_n(p) = k \qquad \text{if } p \in (\tfrac{k}{n+1}, \tfrac{k+1}{n+1}) \text{ (k integer)}$$

If $p = \frac{k}{n+1}$, both k and k-1 can be considered as the mode. The functions $m_n(p)$, $\hat{x}_n(p)$ and np are given in figure 1 for n = 5 and n = 6. The pictures suggest the following theorem.

<u>THEOREM 2</u>. *Let* $p \neq p_i$, $p \neq \frac{i}{n+1}$ *for whole* i, *then for* $p < \frac{1}{2}$

(a) $$m_n(p) = \hat{x}_n(p) \quad \text{or} \quad m_n(p) = \hat{x}_n(p) + 1$$

(b) $$m_n(p) = \hat{x}_n(p) + 1 \Rightarrow \hat{x}_n(p) < np < m_n(p)$$

and for $p > \frac{1}{2}$

(c) $$m_n(p) = \hat{x}_n(p) \quad \text{or} \quad m_n(p) = \hat{x}_n(p) - 1$$

(d) $$m_n(p) = \hat{x}_n(p) - 1 \Rightarrow m_n(p) < np < \hat{x}_n(p).$$

Propositions (a) and (c) can be summarized as

$$0 \leq |\hat{x}_n(p) - \tfrac{1}{2}n| - |m_n(p) - \tfrac{1}{2}n| \leq 1$$

In words the mode is prone to be further from $\frac{1}{2}$ n than the median. It is well-known, that "mean < median < mode" or "mean > median > mode" holds for a large class of continuous distributions. RUNNENBURG [4] gives sufficient conditions. In the binomial case however, if median and mode differ (which they do if e.g. $p_k < p < k/(n+1)$), the mean lies in between, as stated in propositions (b) and (d).

PROOF OF THEOREM 2. Notice, that the first condition provides unique medians and modes. We shall now consider $m_n(p)$, $\hat{x}_n(p)$ and np as functions of p on the interval $(\frac{h}{n+1}, \frac{h+1}{n+1})$ for $0 \le h \le \frac{1}{2}n$. Greater values of h can be dealt with similarly. Let $0 \le h \le \frac{1}{2}n - \frac{1}{2}$, then

$$\hat{x}_n(p) = h \Leftrightarrow p \in (\frac{h}{n+1}, \frac{h+1}{n+1})$$

$$m_n(p) = h \Leftrightarrow p \in (p_h, p_{h+1}) \qquad \text{by (9)}$$

By (10) and (11) we have for $0 \le h \le \frac{1}{2}n - \frac{1}{2}$

$$(15) \qquad p_h \le \frac{h}{n+1} < \frac{h}{n-1} \le p_{h+1} \le \frac{h+1}{n+1}$$

So,

$$p \in (\frac{h}{n+1}, p_{h+1}) \Leftrightarrow \hat{x}_n(p) = m_n(p) = h$$

and

$$p \in (p_{h+1}, \frac{h+1}{n+1}) \Leftrightarrow \hat{x}_n(p) = h \wedge m_n(p) = h + 1$$

By (15)

$$p \in (p_{h+1}, \frac{h+1}{n+1}) \Leftrightarrow h < n\frac{h}{n-1} \le np_{h+1} < np < n\frac{h+1}{n+1} < h + 1$$

If $h = \frac{1}{2}n$, then by (11)

$$p \in (\frac{h}{n+1}, \frac{h+1}{n+1}) \Leftrightarrow \hat{x}_n(p) = m_n(p) = h \qquad \square$$

About the difference between median and mean the following can be said.

<u>THEOREM 3</u>. *For any median* m *(weak or strong)*

$$|np - m| \le \max\{p, 1-p\}.$$

<u>PROOF</u>. Only the case $p < \frac{1}{2}$ is considered. The case $p > \frac{1}{2}$ then follows by symmetry and the case $p = \frac{1}{2}$ is obvious.

If $m_n(p) = m \le \frac{1}{2}n - \frac{1}{2}$ is the strong median, then by (10) or (11)

$$\frac{m-1}{n-1} < p_m < p < p_{m+1} \le \frac{m+1}{n+1} \Rightarrow \tag{16}$$

$$m - np < 1 - p \quad \text{and} \quad np - m < 1 - p \Rightarrow$$

$$|m - np| \le 1 - p = \max\{p, 1-p\}$$

Notice, that the equality sign in (16) holds if and only if $m = \frac{1}{2}n - \frac{1}{2}$.

If $m_n(p) = m = \frac{1}{2}n$ is the strong median, then by (11) and by $p < \frac{1}{2}$

$$\frac{m-1}{n-1} < p_m < p < \frac{1}{2} \Rightarrow \tag{17}$$

$$m - np < 1 - p \quad \text{and} \quad np < \frac{1}{2}n = m \Rightarrow$$

$$0 < m - np < 1 - p = \max\{p, 1-p\}$$

If $[k-1, k]$ is the set of weak medians of the binomial-(n,p) distribution, then by (10) or (11)

$$\frac{k-1}{n-1} < p = p_k < \frac{k}{n+1} \Rightarrow \tag{18}$$

$$k - np < 1 - p \quad \text{and} \quad np - k < -p \Rightarrow$$

$$p < k - np < 1 - p$$

So, for any $m \in [k-1, k]$ we have $p - 1 < m - np < 1 - p$, which completes the proof. $\square$

REFERENCES

[1] JOHNSON, N.L. & S. KOTZ, (1969), *Distributions in Statistics,* Discrete Distributions, Chapter 3, p. 53.

[2] NEUMANN, P., (1966), *Ueber den Median der Binomial - und Poissonverteilung,* Wissenschaftliche Zeitschrift der Technischen Universität Dresden 15, 233-226.

[3] NEUMANN, P., (1970), *Ueber den Median einiger diskreter Verteilungen und eine damit zusammenhängende monotone Konvergenz,* Wissenschaftliche Zeitschrift der Technischen Universität Dresden 19, 29-33.

[4] RUNNENBURG, J.T., (1978), *Mean, median, mode,* Statistica Neerlandica 32, 73-79.

[5] UHLMANN, W., (1963), *Ranggrößen als Schätzfunktionen,* Metrika 7, 23-40.