ADV MATH
SCI JOURNAL

UNIQUENESS OF A MEDIAN OF A BINOMIAL DISTRIBUTION WITH RATIONAL PROBABILITY

Szymon Nowakowski

ABSTRACT. In this paper we show that the median of the binomial distribution $B(n,p)$ is unique for all rational p, with the only exception of $p = \frac{1}{2}$ and n odd.

1. INTRODUCTION

The tail estimates of probabilistic distributions have been the very active research frontier in theoretic (large deviation theory) and applied sciences, both. Not so much work has been devoted to medians, and the existing results seem to be either not that well known or misunderstood. As the example, the famous mean, median and mode inequality theorems [1–4] are a relatively new achievement. The discrete case [4] is the work as recent as the turn of the centuries. Those theorems establish the order of the mean, medians and modes under certain conditions. Nevertheless, there seems to be a widespread belief that the order holds universally, and there exist papers devoted to correcting that belief [5, 6].

Recently, there have been papers that establish the bounds on a distance of a median to the mean in the case of a binomial distribution [7], a binomial and negative binomial distribution [8] and in the case of a binomial and Poisson

distributions [9]. The first of those papers introduces the notion of a *weak* median (a non-unique median) and the *strong* (unique) median. In the same paper sufficient conditions are given for the median to be unique in the binomial distribution. The sufficient conditions are based on a distance between the median candidate and the mean of the binomial distribution.

In this paper we state another sufficient condition of a completely different nature than before (not based on a distance between median candidates and the mean): we show that the median of the binomial distribution $B(n, p)$ is unique (strong in the notation of [7]) for all rational p, with the only exception of $p = \frac{1}{2}$ and n odd.

The main motivation of this paper comes from considering an approximation of a hypergeometric distribution with a binomial distribution, with p being a ratio between certain integer parameters of the hypergeometric distribution being approximated, thus p being rational. In toy problems such as urn problems (drawing with replacement), p is a ratio between integers (numbers of balls). In practical statistical inference, p is frequently a ratio between integer counts, too. In those cases one may apply the results of this paper: that there is the unique median guaranteed to exist (with the only exception of $p = \frac{1}{2}$ and n odd).

2. NOTATION

Binomial distribution $B(n, p)$ with parameters $n \in \mathbb{N}$ and $0 \le p \le 1$ is defined with the use of its discrete probability density function $b(\cdot, n, p)$

$$b(k, n, p) = \binom{n}{k} p^k (1 - p)^{n-k}, \text{ for } k \in \mathbb{N}, \ 0 \le k \le n.$$

Its cumulative distribution function will be denoted $B(\cdot, n, p)$, i.e.,

$$B(k, n, p) = \sum_{i=0}^{k} b(i, n, p), \text{ for } k \in \mathbb{N}, \ 0 \le k \le n.$$

We call a discrete random variable X binomial with a distribution $B(n, p)$ if

$$P(X \le k) = B(k, n, p), \text{ for all } k \in \mathbb{N}, \ 0 \le k \le n.$$

We say that $m \in \mathbb{R}$ is a median of a discretely distributed random variable X if

(2.1) $$P(X \le m) \ge \frac{1}{2} \text{ and } P(X \ge m) \ge \frac{1}{2}.$$

Following [7] we make an additional distinction here: we call $m \in \mathbb{R}$ the unique median if

(2.2) $$P(X \leq m) > \frac{1}{2} \text{ and } P(X \geq m) > \frac{1}{2}.$$

Please note, that we do not need to restrict the last two definitions to binomial distributions only.

3. PREREQUISITES

Before we move to examine the main result of this paper, let us state a few facts to lay the foundation of our understanding of the medians in the discrete distributions.

Lemma 3.1. *If $m \in \mathbb{R}$ is the unique median of a discrete random variable X, then m belongs to the support of X.*

Proof. From the definition of the strong median (2.2) we have

$$P(X > m) = 1 - P(X \leq m) < \frac{1}{2} \text{ and } P(X \geq m) > \frac{1}{2},$$

and thus $P(X = m) > 0$.

$\square$

Lemma 3.2. *If $m \in \mathbb{R}$, X is a discrete random variable then*

(3.1) $$P(X \leq m) = \frac{1}{2} \text{ or } P(X \geq m) = \frac{1}{2},$$

if, and only if, m is a median but not the unique median of X.

Proof. ($\implies$): Suppose that the first equation in (3.1) holds: we have

$$P(X \geq m) \geq P(X > m) = 1 - P(X \leq m) = \frac{1}{2},$$

and thus m is a median by (2.1). By the symmetrical argument for the second equation in (3.1) we can show that m satisfies the definition of a median (2.1) in this case too. Obviously, m fails (2.2) and thus it is not the unique median.

($\impliedby$): Obviously, if (2.1) holds but any one of the two inequalities (2.2) fails, then (3.1) must hold.

This concludes the proof.

$\square$

Lemma 3.3. *If $m \in \mathbb{R}$ is a median but not the unique median of a discrete random variable X with its support being a closed set in $\mathbb{R}$, then there exist uniquely defined $m_1 < m_2$, both in the support of X, such that $m \in [m_1, m_2]$ and all $m' \in [m_1, m_2]$ are medians. Additionally we have*

$$P(X \leq m_1) = P(X \geq m_2) = \frac{1}{2}.$$

Proof. To see that please note, that by the Lemma 3.2 we have

$$P(X \leq m) = \frac{1}{2} \text{ or } P(X \geq m) = \frac{1}{2}.$$

Without losing generality assume that the first equality holds, the argument for the second equality is symmetrical.

If m is in the support of X then let $m_1 = m$. Otherwise, if m is not in the support of X, select $m_1 < m$ from the support of X for which $P(X \leq m_1) = \frac{1}{2}$ still holds. It is possible because the support of X is a closed set in $\mathbb{R}$. Either way, we have found $m_1 \leq m$ with $P(X \leq m_1) = \frac{1}{2}$. Such a point m_1 is uniquely defined.

To find m_2 observe, that we have

$$P(X > m) = 1 - P(X \leq m) = \frac{1}{2},$$

which allows us to find $m_2 > m \geq m_1$, m_2 is from the support of X with the property $P(X \geq m_2) = \frac{1}{2}$. Again, we use the property that the support of X is a closed set in $\mathbb{R}$. Such a point m_2 is uniquely defined, too.

For all $m' \in [m_1, m_2]$ we have

$$P(X \leq m') \geq P(X \leq m_1) = \frac{1}{2} \text{ and } P(X \geq m') \geq P(X \geq m_2) = \frac{1}{2}.$$

Thus $[m_1, m_2]$ is the uniquely defined interval of medians of X. $\qquad\square$

Lemma 3.4. *If $m_1, m_2 \in \mathbb{R}$ are the unique medians of a discrete random variable X, then $m_1 = m_2$.*

Proof. The proof will proceed by showing contradiction. Let us assume that $m_1 \neq m_2$. Without losing generality we can assume that $m_1 < m_2$. The case $m_1 > m_2$ is symmetrical.

By the definition of the unique median (2.2) for m_2 we have

$$P(X \leq m_1) \leq P(X < m_2) = 1 - P(X \geq m_2) < \frac{1}{2}.$$

Thus, m_1 doesn't satisfy (2.2) and thus it is not the unique median, showing contradiction. □

Lemmas 3.3 and 3.4 explain the motivation behind calling a median satisfying (2.2) the *unique* median. Lemmas 3.1 and 3.3 state additionally that this distinction (unique/non-unique) holds in restriction to a support of a discrete random variable with a closed support, too.

4. MAIN RESULT

Shifting the discussion back to binomial distributions, we can see that the discriminating condition for $B(n, p)$ to have a non-unique median is having the value $k - 1$ in the support with $B(k - 1, n, p) = \frac{1}{2}$. Then and only then $k - 1$ and k are the two ends of the interval of non-unique medians. Otherwise, if there is no such $k - 1$ with $B(k - 1, n, p) = \frac{1}{2}$, there is the unique median in $B(n, p)$.

The above observation gives rise to examining, for each $n \geq 1$, the n values $p_{n,1}, \ldots, p_{n,n}$ defined in [7] as follows:

$$(4.1) \qquad B(k - 1, n, p_{n,k}) = \frac{1}{2}, \text{ for } k \in \mathbb{N}, 1 \leq k \leq n.$$

Such $p_{n,k}$ exist, because for fixed k and n, $B(k - 1, n, p)$ is a continuous function of p with values $B(k - 1, n, p = 0) = 1$ and $B(k - 1, n, p = 1) = 0$, thus $\frac{1}{2}$ must be attained by Darboux's theorem at some point $p \in (0, 1)$, too.

In other words, for $1 \leq k \leq n$, $B(n, p_{n,k})$ is the distribution with $k - 1$ and k as the two ends of the interval of non-unique medians.

Theorem 4.1. *Fix $n \geq 1$. The n values $p_{n,1}, \ldots, p_{n,n}$ defined as in (4.1) are irrational, with the only exception $p_{n,\lceil \frac{n}{2} \rceil} = \frac{1}{2}$ for odd n.*

Proof. The proof will proceed in five parts marked with Roman numerals I-V:

I. $p_{n,i} < p_{n,j}$ for $1 \leq i < j \leq n$:

We can see that $n \geq 2$. $B(k, n, p)$ as a binomial CDF is a strictly increasing function of k with n and p fixed and it is a strictly decreasing function of p with n

and k fixed. To see that, fix $n \geq 2$ and $0 \leq j \leq n-1$ and calculate the derivative

$$\frac{d}{dp}B(j,n,p) = \sum_{i=0}^{j}\binom{n}{i}\left(ip^{i-1}(1-p)^{n-i} - p^i(n-i)(1-p)^{n-i-1}\right) =$$

$$= \sum_{i=0}^{j-1}\binom{n}{i+1}(i+1)p^i(1-p)^{n-i-1} - \sum_{i=0}^{j}\binom{n}{i}(n-i)p^i(1-p)^{n-i-1}.$$

It is easy to check that $\binom{n}{i+1}(i+1) = \binom{n}{i}(n-i) = \binom{n-1}{i}n$. It allows telescopic cancelling of respective terms in both sums and we are left only with the last term of the right sum, i.e.,

$$\frac{d}{dp}B(j,n,p) = -n\binom{n-1}{j}p^j(1-p)^{(n-1)-j} = -n \cdot b(j,n-1,p) < 0.$$

For $i < j$ we have

$$B(j-1,n,p_{n,j}) = \frac{1}{2} = B(i-1,n,p_{n,i}) < B(j-1,n,p_{n,i}).$$

So we have $p_{n,i} < p_{n,j}$, which ends this part of the proof.

II. $0 < p_{n,i} = 1 - p_{n,(n-i+1)} < 1$ for $1 \leq i \leq n$:

First note, that if any of $p_{n,i}$ were equal 0 or 1, the distribution would degenerate and its CDF in no point would be equal to $\frac{1}{2}$.

To prove the equation $p_{n,i} = 1 - p_{n,(n-i+1)}$, observe that for random variables X_+ with a binomial distribution $B(n,p_{n,i})$ and X_- with a binomial distribution $B(n,1-p_{n,i})$ we have

$$\frac{1}{2} = B(i-1,n,p_{n,i}) = P(X_+ \leq i-1) = P(X_- \geq n - (i-1)).$$

Thus,

$$B(n-i,n,1-p_{n,i}) = P(X_- \leq n-i) = P(X_- < n-i+1) =$$

$$= 1 - P(X_- \geq n-i+1) = \frac{1}{2} = B(n-i,n,p_{n,(n-i+1)}).$$

It ends this part of the proof.

III. $p_{n,\lceil \frac{n}{2} \rceil} = \frac{1}{2}$ for odd n:

This is a straightforward consequence of part II., after you note that $\lceil \frac{n}{2} \rceil = n - \lceil \frac{n}{2} \rceil + 1$ for odd n.

IV. $p_{n,i}$ is irrational for $i > \lceil \frac{n}{2} \rceil$:

First, observe that for odd n, by using parts I. and III., we have $p_{n,i} > p_{n,\lceil \frac{n}{2} \rceil} = \frac{1}{2}$ for $i > \lceil \frac{n}{2} \rceil$.

For n even, we use parts I. and II. to arrive to the same conclusion: $p_{n,i} > p_{n,(n-i+1)} = 1 - p_{n,i}$, for $i > \frac{n}{2}$, thus $p_{n,i} > \frac{1}{2}$ for $i > \frac{n}{2} = \lceil \frac{n}{2} \rceil$.

By using part II. again, to get the upper bound on $p_{n,i}$, we arrive to the conclusion, that $p_{n,i} \in (\frac{1}{2}, 1)$.

We observe that $p = p_{n,i}$ is the solution to the equation $2B(i - 1, n, p) - 1 = 0$ which is polynomial in relation to p with all coefficients being integers. We examine the coefficient at p^0 after the polynomial $2B(i - 1, n, p) - 1$ is fully expanded:

$$
\begin{aligned}
\left[p^0\right]\Big(2B(i - 1, n, p) - 1\Big) &= \left[p^0\right]\left(2\sum_{j=0}^{i-1}\binom{n}{j}p^j(1 - p)^{n-j}\right) - 1 \\
&= \left[p^0\right]\left(2\binom{n}{0}p^0(1 - p)^n\right) - 1 = 1.
\end{aligned}
$$

If $p_{n,i}$ were rational of the form $\frac{q}{r}$ with q and r relatively prime integers, then (by the Rational Root Theorem) q would divide $\left[p^0\right]\Big(2B(i - 1, n, p) - 1\Big) = 1$. So $p_{n,i}$ would be rational of the form $\frac{1}{r}$. But that is impossible, because in $(\frac{1}{2}, 1)$ interval there are no rational numbers of the form $\frac{1}{r}$ with integer r.

This finalizes this part of the proof.

V. $p_{n,i}$ is irrational for $i \neq \lceil \frac{n}{2} \rceil$ or n even.

This is a straightforward consequence of parts II. and IV. It also finalizes the proof of the main result of this paper. $\qquad\square$

REFERENCES

[1] J.TH. RUNNENBURG: *Mean, median, mode*, Stat. Neerl., **32**(2) (1978), 73–79.

[2] W.R. VAN ZWET: *Mean, median, mode II*, Stat. Neerl., **33**(1) (1979), 1–5.

[3] S.W. DHARMADHIKARI, K. JOAG-DEV: *Mean, median, mode III*, Stat. Neerl., **37**(4) (1983), 165–168.

[4] B. ABDOUS, R. THEODORESCU: *Mean, median, mode IV*, Stat. Neerl., **52**(3) (1998), 356–359.

[5] K.M. ABADIR: *The Mean-Median-Mode Inequality: Counterexamples*, Econometric Theory, **21**(2) (2005), 477–482.

[6] P.T. VON HIPPEL: *Mean, Median, and Skew: Correcting a Textbook Rule*, J. Stat. Educ., **13**(2) (2005).

[7] R. KAAS, J.M. BUHRMAN: *Mean, Median and Mode in Binomial Distributions*, Stat. Neerl., **34**(1) (1980), 13–18.

[8] R. GÖB: *Bounds for median and 50 percentage point of binomial and negative binomial distribution*, Metrika, **41**(1) (1994), 43–54.

[9] K. HAMZA: *The smallest uniform upper bound on the distance between the mean and the median of the binomial and Poisson distributions*, Stat. Probab. Lett., **23**(1) (1995), 21–25.

AI INVESTMENTS
OKRZEI 2, SKIERNIEWICE
POLAND.
Email address: szymon.nowakowski@aiinvestments.pl