The sample monomode and an associated test for discrete monomodality

Raúl Pérez-Fernández & Bernard De Baets

To cite this article: Raúl Pérez-Fernández & Bernard De Baets (2019) The sample monomode and an associated test for discrete monomodality, Communications in Statistics - Theory and Methods, 48:21, 5419-5426, DOI: 10.1080/03610926.2018.1513148

To link to this article: https://doi.org/10.1080/03610926.2018.1513148

Published online: 25 Jan 2019.
The sample monomode and an associated test for discrete monomodality

Raúl Pérez-Fernández and Bernard De Baets

KERMIT, Department of Data Analysis and Mathematical Modelling, Ghent University, Gent, Belgium.

ABSTRACT

The notions of (sample) mean, median and mode are common tools for describing the central tendency of a given probability distribution. In this article, we propose a new measure of central tendency, the sample monomode, which is related to the notion of sample mode. We also illustrate the computation of the sample monomode and propose a statistical test for discrete monomodality based on the likelihood ratio statistic.

ARTICLE HISTORY

Received 5 December 2017
Accepted 14 August 2018

KEYWORDS

Discrete unimodality; discrete monomodality; likelihood ratio; multinomial distribution; monomode

1. Introduction

Unimodal distributions have been largely studied in the context of continuous probability distributions (Khinchine 1938; Dharmadhikari and Jogdeo 1976). There exist several tests for the unimodality of a continuous probability distribution, Silverman’s test (1981, 1983) (see also (Hall and York 2001)) and the DIP test (Hartigan and Hartigan 1985) probably being the most prominent ones. In the discrete case, the literature is sparser, and we commonly find two different definitions of unimodality: (a) the mode is unique (Dutta and Goswami 2010) and (b) the probabilities are increasing on the left side of the mode(s) and decreasing on the right side of the mode(s) (Medgyessy 1972; Bertin and Theodorescu 1984; Ushakov 1998). Since the second definition has interesting implications for defining a natural measure of central tendency (Pérez-Fernández, Sader, and De Baets 2018), we consider the latter and use the term monomodality instead in order to avoid any potential confusion.

In this article, we propose a new measure of central tendency, the sample monomode, and a new test for discrete monomodality based on an existing likelihood ratio test for multinomial distributions under inequality constraints (El Barmi and Dykstra 1995; Davis-Stober 2009).

2. The (sample) monomode

Let \( \mathcal{L} = \{L_1, \ldots, L_k\} \) be a finite linearly ordered set \((L_i \leq \mathcal{L} L_j \text{ if } i \leq j)\) and \( p \) be a probability mass function on \( \mathcal{L} \). An element \( L \in \mathcal{L} \) is called a mode of \( p \) if, for any \( L' \in \mathcal{L} \), it holds that \( p(L') \leq p(L) \) (Dutta and Goswami 2010). An element \( L \in \mathcal{L} \) is called...
a monomode of \( p \) if, for any \( L', L'' \in \mathcal{L} \) satisfying that \( L \leq \varphi L' \leq \varphi L'' \) or \( L'' \leq \varphi L' \leq \varphi L \), it holds that \( p(L'') \leq p(L') \leq p(L) \). Obviously, any monomode is a mode but the converse does not hold. Unlike a mode, a monomode is not assured to exist. Like a mode, a monomode is not assured to be unique (in case it exists).

A probability mass function on \( \mathcal{L} \) is said to be (discrete) monomodal if a monomode exists. Equivalently, a probability mass function is monomodal if it is increasing on the left side of the mode(s) and decreasing on the right side of the mode(s).\(^1\) Note that monomodality is a common point of discussion in statistics and probability theory, usually referred to as unimodality (Medgyessy 1972; Bertin and Theodorescu 1984; Ushakov 1998).

Consider a random sample (of size \( n \)) from \( p \). The vector \((x_1, \ldots, x_k)\) in which each \( x_i \) represents the number of times element \( L_i \) appears in the sample, can then be understood as an observation of a random variable \( X = (X_1, \ldots, X_k) \) that follows a multinomial distribution with parameters \((n, p)\), where \( n \) is determined by the sample size and \( p = (p(L_1), \ldots, p(L_k)) \). Under the assumption of monomodality, we want to identify the sample monomode(s), i.e., the element(s) in \( \mathcal{L} \) that is(are) the most likely to be the monomode(s) given the sample \((x_1, \ldots, x_k)\).

We recall that the probability mass function of a multinomial distribution is given by:

\[
p(x_1, \ldots, x_k) = \frac{n!}{x_1! \cdots x_k!} \prod_{i=1}^{k} p_i^{x_i}
\]

where \( x_1, \ldots, x_k \) are non negative integers such that \( \sum_{i=1}^{k} x_i = n \).

The associated likelihood function \( L(p; x_1, \ldots, x_k) \) is proportional to \( \prod_{i=1}^{k} p_i^{x_i} \). It is known that, for a multinomial model in which no constraints are imposed on the vector of parameters \( p \), the maximum likelihood estimator \( \hat{p} \) of \( p \) is given by:

\[
\hat{p} = \left( \frac{x_1}{n}, \ldots, \frac{x_k}{n} \right)
\]

However, under the assumption of monomodality, we would expect the values \( p_l \) to satisfy the inequality constraints associated with a monomodal distribution. Note that the parameter space for \( p \) will now be restricted to

\[
\Theta = \bigcup_{\ell=1}^{k} \Theta_{\ell}
\]

where \( \Theta_{\ell} \) is the parameter space in case monomodality holds w.r.t. \( L_{\ell} \):

\[
\Theta_{\ell} = \left\{ p \in [0, 1]^k \left| \sum_{i=1}^{k} p_i = 1 \right. \wedge \left. (i \leq j \leq \ell \forall \ell \leq j \leq i) \Rightarrow p_i \leq p_j \leq p_\ell \right. \right\}
\]

Note that the dimension of each \( \Theta_{\ell} \) (and thus of \( \Theta \)) is \( k - 1 \).

\(^1\)In case the mode is not unique, all the modes need to be consecutive elements of \( \leq \varphi \).
Some studies on the incorporation of inequality constraints in a multinomial setting (Davis-Stober 2009) have proven that, for each $\ell$, a unique maximum likelihood estimator $\hat{p}_\ell$ of $p$ in $\Theta_\ell$ exists, and, in case $\hat{p}_\ell$ violates at least one of the inequality constraints, lies on the boundary of the corresponding $\Theta_\ell$. Thus, a maximum likelihood estimator $\hat{p}$ of $p$ in $\Theta$ is always assured to exist, and, although it might not be unique, there are at most $k$ of them (one per $\Theta_\ell$).

Thus, we propose a new measure of central tendency of a sample—the sample monomode(s)—which is(are) the element(s) $L_\ell$ such that $\hat{p}_\ell$ is a maximum likelihood estimator of $p$ in $H_\ell$. Obviously, if the (absolute) frequencies associated with the sample $(x_1, ..., x_k)$ are increasing on the left side of $x_\ell$ and decreasing on the right side of $x_\ell$, then $L_\ell$ is a sample monomode, and, if in addition the mode is unique, then $L_\ell$ is the unique sample monomode.

**Example 2.1.** Consider the sample $(3, 1, 9, 15, 40, 20, 5, 6, 1)$. Although $L_5$ is the sample mode, the frequencies neither increase on its left side, nor decrease on its right side: element $L_1$ appears three times whereas element $L_2$ appears once, and element $L_7$ appears five times whereas element $L_8$ appears six times. This is illustrated in the left side of Figure 1.

For each $\ell \in \{1, ..., k\}$, we compute $\hat{p}_\ell$ by minimizing $L(p) = -2 \log L(p; x_1, ..., x_k)$ (note that a minimizer of $L(p)$ is a maximizer of $L(p; x_1, ..., x_k)$). All these vectors are gathered in Table 1. We can see that the (unique) maximum likelihood estimator of $p$ in $\Theta$ is $\hat{p} = \hat{p}_5$. Thus, $L_5$ is the sample monomode.

**Figure 1.** Representation of the samples in Example 2.1 (left) and in Example 2.2 (right).

**Table 1.** Vectors $\hat{p}_\ell$ and $p$, and associated $-2 \log$-likelihood.

<table>
<thead>
<tr>
<th>$\ell$</th>
<th>$\hat{p}_\ell$</th>
<th>$L(\hat{p}_\ell)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$(0.1467, 0.1467, 0.1467, 0.1467, 0.1467, 0.1467, 0.0550, 0.0550, 0.0100)$</td>
<td>410.8680</td>
</tr>
<tr>
<td>2</td>
<td>$(0.0300, 0.1700, 0.1700, 0.1700, 0.1700, 0.0550, 0.0550, 0.0100)$</td>
<td>395.2917</td>
</tr>
<tr>
<td>3</td>
<td>$(0.0200, 0.0200, 0.2133, 0.2133, 0.2133, 0.2000, 0.0550, 0.0550, 0.0100)$</td>
<td>366.4405</td>
</tr>
<tr>
<td>4</td>
<td>$(0.0200, 0.0200, 0.0900, 0.2750, 0.2750, 0.2000, 0.0550, 0.0550, 0.0100)$</td>
<td>354.0446</td>
</tr>
<tr>
<td>5</td>
<td>$(0.0200, 0.0200, 0.0900, 0.1500, 0.4000, 0.2000, 0.0550, 0.0550, 0.0100)$</td>
<td>342.2532</td>
</tr>
<tr>
<td>6</td>
<td>$(0.0200, 0.0200, 0.0900, 0.1500, 0.3000, 0.3000, 0.0550, 0.0550, 0.0100)$</td>
<td>349.0492</td>
</tr>
<tr>
<td>7</td>
<td>$(0.0200, 0.0200, 0.0900, 0.1500, 0.2167, 0.2167, 0.2167, 0.0600, 0.0100)$</td>
<td>373.3455</td>
</tr>
<tr>
<td>8</td>
<td>$(0.0200, 0.0200, 0.0900, 0.1500, 0.1775, 0.1775, 0.1775, 0.1755, 0.0100)$</td>
<td>386.2506</td>
</tr>
<tr>
<td>9</td>
<td>$(0.0200, 0.0200, 0.0900, 0.1500, 0.1450, 0.1450, 0.1450, 0.1450, 0.1450)$</td>
<td>410.6370</td>
</tr>
<tr>
<td>$p$</td>
<td>$(0.0300, 0.0100, 0.0900, 0.1500, 0.4000, 0.2000, 0.0550, 0.0600, 0.0100)$</td>
<td>341.1157</td>
</tr>
</tbody>
</table>
Table 2. Vectors $\hat{p}_\ell$ and $p$, and associated $-2 \log$-likelihood.

<table>
<thead>
<tr>
<th>$\ell$</th>
<th>$\hat{p}_\ell$</th>
<th>$L(\hat{p}_\ell)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(0.3700, 0.2500, 0.0543, 0.0543, 0.0543, 0.0543, 0.0543, 0.0543, 0.0543, 0.0543)</td>
<td>364.3149</td>
</tr>
<tr>
<td>2</td>
<td>(0.3100, 0.3100, 0.0543, 0.0543, 0.0543, 0.0543, 0.0543, 0.0543, 0.0543, 0.0543)</td>
<td>366.6523</td>
</tr>
<tr>
<td>3</td>
<td>(0.2067, 0.2067, 0.2067, 0.0633, 0.0633, 0.0633, 0.0633, 0.0633, 0.0633, 0.0633)</td>
<td>405.2145</td>
</tr>
<tr>
<td>4</td>
<td>(0.1550, 0.1550, 0.1550, 0.1550, 0.0760, 0.0760, 0.0760, 0.0760, 0.0760, 0.0760)</td>
<td>427.0306</td>
</tr>
<tr>
<td>5</td>
<td>(0.1240, 0.1240, 0.1240, 0.1240, 0.1240, 0.1240, 0.0950, 0.0950, 0.0950, 0.0950)</td>
<td>437.7415</td>
</tr>
<tr>
<td>6</td>
<td>(0.1111, 0.1111, 0.1111, 0.1111, 0.1111, 0.1111, 0.1111, 0.1111, 0.1111, 0.1111)</td>
<td>439.4481</td>
</tr>
<tr>
<td>7</td>
<td>(0.1033, 0.1033, 0.1033, 0.1033, 0.1033, 0.1033, 0.1267, 0.1267, 0.1267, 0.1267)</td>
<td>438.4855</td>
</tr>
<tr>
<td>8</td>
<td>(0.0886, 0.0886, 0.0886, 0.0886, 0.0886, 0.0886, 0.0886, 0.1900, 0.1900, 0.1900)</td>
<td>426.7849</td>
</tr>
<tr>
<td>9</td>
<td>(0.0775, 0.0775, 0.0775, 0.0775, 0.0775, 0.0775, 0.0775, 0.0775, 0.0775, 0.0775)</td>
<td>390.6636</td>
</tr>
<tr>
<td>$p$</td>
<td>(0.3700, 0.2500, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000)</td>
<td>216.4258</td>
</tr>
</tbody>
</table>

Example 2.2. Consider the sample $\{37, 25, 0, 0, 0, 0, 0, 38\}$. Although $L_0$ is the sample mode, the frequencies do not increase on its left side: elements $L_1$ and $L_2$ are the only other elements appearing in the sample. This is illustrated in the right side of Figure 1.

For each $\ell \in \{1, ..., k\}$, we compute $\hat{p}_\ell$. All these vectors are gathered in Table 2. We can see that the (unique) maximum likelihood estimator of $p$ in $\Theta$ is $\hat{p} = \hat{p}_1$. Thus, $L_1$ is the sample monomode. Note that the sample mode and the sample monomode do not coincide.

3. A new test for discrete monomodality

In the preceding section, we have introduced the notion of a sample monomode. In this section, we provide a statistical test for determining whether the underlying discrete probability distribution is monomodal or not given the sample monomode. For an interesting review on hypothesis testing under inequality constraints for multinomial models, we refer to Davis-Stober (2009).

Formally, our hypothesis test is the following:

$$H_0 : p \in \Theta$$
$$H_1 : p \notin \Theta$$

Since $\Theta$ is defined as the union of convex subsets of $[0, 1]^k$, we need to split the hypothesis test into $k$ different hypothesis tests, one for each convex subset. Then the above hypothesis test is rejected if and only if all the following hypothesis tests are rejected ($\ell \in \{1, ..., k\}$) (Berger 1982):

$$H_0^\ell : p \in \Theta_\ell$$
$$H_1^\ell : p \notin \Theta_\ell$$

In case $\Theta_\ell$ would be the result of restricting $[0, 1]^k$ with some equality constraints, Wilk’s theorem (1938) states that the asymptotic distribution of the likelihood ratio test statistic

$$\Lambda(\hat{p}_\ell) = L(\hat{p}_\ell) - L(p)$$

is a chi-squared distribution with the number of degrees of freedom equal to the difference between the cardinalities of the original parameter space and $\Theta_\ell$.

Unfortunately, this is not the case when we are dealing with inequality constraints. Now, the asymptotic distribution of the likelihood ratio test statistic is a chi-bar squared
distribution (Shapiro 1985), which is the average of independent chi-squared distributions with different degrees of freedom.

Let \( X_0, X_1, \ldots, X_k \) be \( k+1 \) independent random variables such that \( X_i \) follows a \( \chi^2_i \) (for any \( i \)) and \( \mathbf{w} \in [0, 1]^{k+1} \) is a weight vector such that \( \sum_{i=0}^{k} w_i = 1 \). The distribution of the random variable \( X = \sum_{i=0}^{k} w_i X_i \) is called a chi-bar squared distribution and is denoted by \( \bar{\chi}^2_w \).

Many studies have dealt with the identification of the weights of a chi-bar squared distribution. We highlight the study addressed in (El Barmi and Dykstra 1995) in which an analytical expression of the weights is given. Unfortunately, as discussed in a later article by one of the authors (El Barmi and Johnson 2006), these weights “do not exist in general in a closed form.” Following the detailed description given in (Davis-Stober 2009), we used Silvapulle’s algorithm (1996) for estimating these weights.

Once the vector of weights \( \mathbf{w} \) has been identified (or estimated), we can construct the rejection region, which is formed by all the instances resulting in large values of \( K(b_{p'}) \).

The rejection region of the test \( H_0^5 \) against \( H_1^5 \) is of the following form:

\[
S^5 = \sum_{i=0}^{k} w_i \left( \chi^2_i \right)^{-1}(\alpha), +\infty[
\]

where \( \alpha \) is the chosen significance level.

**Example 3.1.** Continue with Example 2.1. We recall that \( \hat{\mathbf{p}} = \hat{\mathbf{p}}_5 \). Therefore, we first consider the test

\[
\begin{align*}
H_0^5 & : \mathbf{p} \in \Theta_5 \\
H_1^5 & : \mathbf{p} \not\in \Theta_5
\end{align*}
\]

The associated value of the likelihood ratio test statistic is

\[
\Lambda(\hat{\mathbf{p}}_5) = \mathcal{L}(\hat{\mathbf{p}}_5) - \mathcal{L}(\hat{\mathbf{p}}) = 342.2532 - 341.1157 = 1.1375
\]

The estimation of the vector of weights for this test results in the following vector

\[
\mathbf{w} = (0.2181, 0.4880, 0.2939, 0, 0, 0, 0, 0, 0)
\]

For a significance level of \( \alpha = 0.05 \), the corresponding values of \( (\chi^2_i)^{-1}(\alpha) \) are given by

\[
(0, 3.8415, 5.9915, 7.8147, 9.4877, 11.0705, 12.5916, 14.0671, 15.5073, 16.09190)
\]

Thus, the rejection region for testing \( H_0^5 \) against \( H_1^5 \) is given by:

\[
S^5 = [3.6355, +\infty[
\]

Since it holds that \( \Lambda(\hat{\mathbf{p}}_5) = 1.1375 \not\in S^5 \), we do not reject \( H_0^5 \), and, thus, we do not reject the monomodality hypothesis \( H_0 \).

This result hints that the sample monomode—\( L_5 \)—is a natural measure of central tendency for the given sample.

---

\(^2\)A \( \chi^2_0 \) distribution is a degenerate distribution localized at 0.
**Example 3.2.** Continue with Example 2.2. We recall that \( \hat{\mathbf{p}} = \hat{\mathbf{p}}_1 \). Therefore, we first consider the test

\[
H_0^1 : \mathbf{p} \in \Theta_1 \\
H_1^1 : \mathbf{p} \notin \Theta_1
\]

The associated value of the likelihood ratio test statistic is

\[
\Lambda(\hat{\mathbf{p}}_1) = \mathcal{L}(\hat{\mathbf{p}}_1) - \mathcal{L}(\hat{\mathbf{p}}) = 364.3149 - 216.4258 = 147.8891
\]

The estimation of the vector of weights for this test results in the following vector

\[
\mathbf{w} = (0.0018, 0.0226, 0.1108, 0.2702, 0.3302, 0.2119, 0.0525, 0, 0, 0)
\]

Obviously, the corresponding values of \((x^2_1)^{-1}(x)\) for a significance level of \(\alpha = 0.05\) coincide with those listed in Example 3.1. Thus, the rejection region for testing \(H_0^1\) against \(H_1^1\) is given by:

\[
S^1 = [9.0020, +\infty[
\]

Since it holds that \(\Lambda(\hat{\mathbf{p}}_1) = 147.8891 \in S^1\), we reject \(H_0^1\).

Similarly, we verify that

\[
\begin{align*}
\Lambda(\hat{\mathbf{p}}_2) & = 150.2265 \in S^2 = [10.1619, +\infty[ \\
\Lambda(\hat{\mathbf{p}}_3) & = 188.7887 \in S^3 = [10.1333, +\infty[ \\
\Lambda(\hat{\mathbf{p}}_4) & = 210.6049 \in S^4 = [10.0938, +\infty[ \\
\Lambda(\hat{\mathbf{p}}_5) & = 221.3157 \in S^5 = [10.2406, +\infty[ \\
\Lambda(\hat{\mathbf{p}}_6) & = 223.0223 \in S^6 = [12.0710, +\infty[ \\
\Lambda(\hat{\mathbf{p}}_7) & = 222.0598 \in S^7 = [10.9587, +\infty[ \\
\Lambda(\hat{\mathbf{p}}_8) & = 210.3591 \in S^8 = [11.1972, +\infty[ \\
\Lambda(\hat{\mathbf{p}}_9) & = 174.2378 \in S^9 = [10.3796, +\infty[ \\
\Lambda(\hat{\mathbf{p}}_{10}) & = 174.2378 \in S^{10} = [10.3796, +\infty[ 
\end{align*}
\]

We conclude that we reject \(H_0^\ell\) for all \(\ell \in \{1, \ldots, 9\}\). Thus, we reject the monomodality hypothesis \(H_0^\ell\).

This result hints that the sample monomode—\(L_1\)—is not a natural measure of central tendency for the given sample since the given probability distribution seems not to be monomodal.

**4. Case study**

We now consider a real-life dataset coming from the popular *Internet Movie Database*—**IMDb** (accessed 17/05/2018). Said database gathers millions of fan reviews and ratings. We restrict our attention to ratings provided by underage users; otherwise, due to the large number of reviews available, the rejection of the test basically depends on whether the given sample is monomodal or not. Note that the linearly ordered set \(L = \{L_1, \ldots, L_{10}\}\) used by IMDb users consists of ten elements (i.e., \(k = 10\)).

We first consider the critically acclaimed movie: *The Lion King* (1994). The corresponding frequencies are given by \((8, 8, 5, 4, 33, 83, 225, 502, 521, 605)\) and are illustrated in the top left side of Figure 2. The sample monomode is \(L_{10}\) and the monomodality
hypothesis is not rejected (considering the significance level of $\alpha = 0.05$), thus hinting that $L_{10}$ is a natural measure of the central tendency for this movie. Secondly, we consider a movie considered among the most polarizing movies of all time: *The Room* (2003). The corresponding frequencies are given by $(126, 46, 29, 21, 17, 12, 10, 7, 161)$ and are illustrated in the top right side of Figure 2. The sample monomode can be verified to be again $L_{10}$ but the monomodality hypothesis is now rejected, thus hinting that $L_{10}$ is not a natural measure of the central tendency for this movie. We could conclude that this movie is indeed quite polarizing.

We now consider two other movies that are widely considered to be polarizing: *Titanic* (1997) and *Avatar* (2009). The corresponding frequencies are given by $(24, 5, 15, 19, 46, 119, 303, 449, 275, 544)$ for the former and by $(20, 15, 15, 30, 77, 167, 354, 470, 271, 294)$ for the latter. Unlike the previous two movies, just by looking at the relative frequencies in Figure 2, it is not easy to predict whether the monomodality hypothesis would be rejected or not. Performing the test leads to a rejection in the case of *Titanic* (1997) and to a non rejection in the case of *Avatar* (2009), the sample monomode being $L_{10}$ and $L_8$, respectively.

### 5. Discussion

We have presented a new test for discrete monomodality based on an existing likelihood ratio test for multinomial distributions under inequality constraints. This test relies on the computation of the sample monomode, which is a newly introduced measure of central tendency that coincides with the sample mode if the frequencies are increasing on the left side of the sample mode(s) and decreasing on the right side of the sample mode(s). Ultimately, the proposed test can also be used for deciding
whether the sample monomode is an interesting measure of central tendency or not for the given sample. Further research concerns the study of the power of the test.

**Funding**

Raúl Pérez-Fernández acknowledges the support of the Research Foundation of Flanders (FWO17/PDO/160).

**References**


