Observed Data Containing One Parameter and Chance Error: Probabilistic Evaluation of Parameter by Pythagorean Mean

Dhritikesh Chakrabarty
Department of Statistics, Handique Girls’ College, Guwahati, Assam, India, dhritikesh.c@rediffmail.com, dhritikeshchakrabarty@gmail.com

Abstract

Recently some methods have been developed for determining the appropriate value of the parameter from observed data containing the parameter itself and chance error since the existing statistical methods of estimation in such situation fail in finding out the appropriate value of the parameter. However, the methods are based on some probabilistic assumptions. Accordingly, the value of the parameter obtained by the methods is not deterministic but probabilistic i.e. one cannot be fully certain that the value of the parameter obtained is identical with its actual value. This paper is based on the evaluation of the probability that the value obtained is the actual value of the parameter and on one numerical application of the method in determining the central tendency of each of annual maximum and annual minimum of ambient air temperature at Guwahati.

Key words: 1; Observed data 2; parameter 3; chance error 4; probabilistic evaluation of parameter

1. Introduction:

There are many situations where observed data

\[x_1 , x_2 , \ldots , x_n\]

are composed of some parameter \(\mu\) and chance errors \(\varepsilon_i\) i.e.

\[x_i = \mu + \varepsilon_i , \quad (i = 1 , 2 , \ldots , n)\]


provides $\bar{X}$ as estimator of the parameter $\mu$ where $\bar{X}$ is given by

$$X = \frac{1}{n} \sum_{i=1}^{n} x_i \quad \text{------------------------------------} (1.2)$$

It has been shown that this estimator $\bar{X}$ of the parameter $\mu$ suffers from an error $\bar{E}$ given by

$$\bar{E} = \frac{1}{n} \sum_{i=1}^{n} e_i \quad \text{------------------------------------} (1.3)$$

which is not zero usually. In other words, none of these methods can provide appropriate value of the parameter $\mu$ [Chakrabarty (2014a, 2014b, 2014c)].

Recently, some studies have been done on determining the true value of the parameter $\mu$ involved in the model described by (1.1) [Chakrabarty (2014a, 2014b, 2014c, 2015a, 2015b, 2015c, 2015d, 2015e, 2015f, 2016a, 2016b), Bordoloi & Chakrabarty (2015, 2015–16, 2016a, 2016b, 2016c, 2016–17)]. In these studies some methods have been developed for determining the true value of the parameter $\mu$ when $e_i$ occurs due to chance only. One of them is based on computing sequence of interval value of $\mu$ with decreasing length of interval and then to find out the shortest interval value of $\mu$ [Chakrabarty (2014a, 2014b, 2014c), 2015d] , Bordoloi & Chakrabarty (2016a, 2016b). The other one is based on stable mid range and median. However, these methods may not be always successful in determining the true value $\mu$ [Chakrabarty (2015b), Bordoloi & Chakrabarty (2015)]. For this reason, another method has been derived for determining the true value $\mu$ which is based on the convergence of statistic i.e. some function of the available numerical data (Chakrabarty, 2017). However, in some situations, the available data may not be sufficient obtaining the converging point of the statistic considered. One method has been developed for determining the true value $\mu$ in such situation (Chakrabarty, 2018a). The method is based on the Pythagorean means [Kolmogorov (1930), O'Meara (1989), Riedweg (2005), Cornelli, McKirahan & Macris (2013), de Carvalho (2016), Chakrabarty (2018b, 2018c, 2018d, 2018e)]. The methods, developed so far, for determining the appropriate value of the parameter from observed data containing the parameter itself and chance error are based on some probabilistic assumptions. Accordingly, the value of the parameter obtained by the methods is not
deterministic but probabilistic i.e. one cannot be fully certain that the value of the parameter obtained is identical with its actual value. This paper is based on the evaluation of the probability that the value obtained is the actual value of the parameter and on one numerical application of the method in determining the central tendency of each of annual maximum and annual minimum of ambient air temperature at Guwahati.

2. Probabilistic Evaluation of \( \mu \):

2.1. Probabilistic Convergence of Pythagorean Arithmetic Mean of Observations

If the observations

\[
x_1, x_2, \ldots, x_n
\]

are composed of some parameter \( \mu \) and chance errors then the observations can be expressed as

\[
x_i = \mu + \varepsilon_i, \quad (i = 1, 2, \ldots, n)
\]

where

(i) \( x_1, x_2, \ldots, x_n \) are observed data,

(ii) \( \mu \) is the parameter

& (iii) \( \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n \) are the chance errors associated to

\[
x_1, x_2, \ldots, x_n
\]

respectively which assume positive and negative values in random order.

From Equations described by (2.1),

\[
\sum_{i=1}^{n} x_i = n\mu + \sum_{i=1}^{n} \varepsilon_i
\]

which implies

\[
A(x_1, x_2, \ldots, x_n) = \mu + A(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)
\]

where

\[
A(x_1, x_2, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^{n} x_i
\]

\[
& A(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i
\]

Now,

\[
\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n
\]
are independently and identically distributed random variables with arithmetic expectation zero (0).

Therefore by the law of large numbers, the series

\[ \{ A_n(\varepsilon) = A(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i \} \]

converges to 0 with probability approaching 1 that is with probability approaching certainity as \( n \to \infty \).

Accordingly with probability approaching 1 that is with probability approaching certainity, the series

\[ \{ A_n(x) = A(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) = \frac{1}{n} \sum_{i=1}^{n} x_i \} \]

converges to \( \mu \) as \( n \to \infty \).

**Note (2.1.1):** Since

\[ \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n \]

are random, each of them can assume either positive value or negative value with equal probability i.e.

\[ P(\varepsilon_i < 0) = P(\varepsilon_i > 0) = \frac{1}{2} \quad \text{for all} \quad i = 1, 2, \ldots, n \]

This implies, the probability that all of them assume positive values is \((\frac{1}{2})^n\).

Similarly, the probability that all of them assume negative values is \((\frac{1}{2})^n\).

Therefore, the probability that all of them assume values with same sign is \((\frac{1}{2})^n + (\frac{1}{2})^n = (\frac{1}{2})^{n-1}\).

The series \( \{ A_n(\varepsilon) \} \) will never converge to 0 if all

\[ \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n \]

assume values with same sign

and hence the series \( \{ A_n(x) \} \) will never converge to \( \mu \) with probability \((\frac{1}{2})^{n-1}\) i.e. the converging point of the series \( \{ A_n(x) \} \), if exits, is not the actual value of \( \mu \) with probability \((\frac{1}{2})^{n-1}\).

**2.2. Probabilistic Convergence of Pythagorean Geometric Mean of Observations**

If the observations

\[ x_1, x_2, \ldots, x_n \]

are composed of some parameter \( \mu \) and chance errors then the observations can also be expressed as

\[ x_i = \mu \cdot e_i \quad (i = 1, 2, \ldots, n) \]

where
are the chance errors associated to

respectively assuming positive values either pure decimal fraction or greater than 1 occurred in random order.

From Equations described by (2.6),

which implies

where

Now,

are independently and identically distributed random variables with geometric expectation one (1). Therefore by the similar logic of the law of large numbers, the series

converges to 1 with probability approaching 1 as \( n \to \infty \).

Accordingly, with probability approaching certainty, the series

converges to \( \mu \) as \( n \to \infty \).

Note (2.2.1): Since

are random, each of them can assume positive values either pure decimal fraction and greater than 1 with equal probability i.e.

for all \( i = 1, 2, \ldots, n \).
This implies, the probability that all of them assume values in \( 0 < e_i < 1 \) is \( (\frac{1}{2})^n \).

Similarly, the probability that all of them assume values in \( e_i > 1 \) is \( (\frac{1}{2})^n \).

Therefore, the probability that all of them assume values of same type is \( (\frac{1}{2})^n + (\frac{1}{2})^n = (\frac{1}{2})^{n-1} \).

Now, the series \( \{ G_n(e) \} \) will never converge to 1 if all \( e_1, e_2, \ldots, e_n \) assume values of same type and hence the series \( \{ G_n(e) \} \) will never converge to \( \mu \) with probability \( (\frac{1}{2})^{n-1} \) i.e. the converging point of the series \( \{ G_n(e) \} \), if exits, is not the actual value of \( \mu \) with probability \( (\frac{1}{2})^{n-1} \).

### 2.3. Evaluation of \( \mu \)

Each of the two series given by (2.5) & (2.10) converges to \( \mu \) as \( n \to \infty \) with probability approaching 1 (i.e. with probability approaching certainty).

Therefore, in order to determine the value of \( \mu \), it is required to compute the converging values of the two series

\[
\{ A_n(x) = A(e_1, e_2, \ldots, e_n) = \frac{1}{n} \sum_{i=1}^{n} x_i \}
\]

\[
\& \quad \{ G_n(x) = G(x_1, x_2, \ldots, x_n) = (\prod_{i=1}^{n} x_i)^{1/n} \}
\]

The common value of them is the actual value of \( \mu \) with probability approaching 1 (i.e. with probability approaching certainty) and is not the actual value of \( \mu \) with probability \( (\frac{1}{2})^{n-1} \) where \( n \) is the number of observations.

**Note (2.3.1):** If the series is found to converge but fail to yield a common converging point for the available data then it is to be understood that the available data are insufficient for obtaining the value of \( \mu \).

**Note (2.3.2):** If the series is found either not to converge or to converge to different points then it is to be understood that the errors involved in the data are not only due to chance but also due to some assignable cause(s). Consequently, the data do not follow the model described by the equation (2.1). Accordingly, the value of \( \mu \) cannot be determined from the given data in this case.

**Note (2.3.3):** From (2.6),
\[
\log G(x_1, x_2, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^{n} \log x_i
\]

which yields,
\[
G(x_1, x_2, \ldots, x_n) = \text{antilog} \{ \frac{1}{n} \sum_{i=1}^{n} \log x_i \} \quad \text{----------------------------------- (2.11)}
\]

This formula can be applied in computing the values of the series given by (2.10) since its computation by
\[
G(x_1, x_2, \ldots, x_n) = (\prod_{i=1}^{n} x_i)^{1/n}
\]

is too complicated.

3. Application to Numerical Data:

Observed data considered here are the data on each of annual maximum & annual minimum of ambient air temperature, occurred in temperature periodic year (TPR), at Guwahati during the period from 1969 to 2013. The objective here is to evaluate, with probability approaching certainty, the central tendency of each of annual maximum & annual minimum of ambient air temperature at Guwahati

3.1. Annual Maximum of Ambient Air Temperature at Guwahati:

The following table shows the observed data on annual maximum of ambient air temperature, occurred in temperature periodic year (TPR), at Guwahati during the period from 1969 to 2013:

<table>
<thead>
<tr>
<th>TPR No ((i))</th>
<th>Observed Value ((x_i))</th>
<th>TPR No ((i))</th>
<th>Observed Value ((X_i))</th>
<th>TPR No ((i))</th>
<th>Observed Value ((x_i))</th>
<th>TPR No ((i))</th>
<th>Observed Value ((X_i))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>37.1</td>
<td>12</td>
<td>35.1</td>
<td>23</td>
<td>37.4</td>
<td>34</td>
<td>38.0</td>
</tr>
<tr>
<td>2</td>
<td>36.6</td>
<td>13</td>
<td>35.8</td>
<td>24</td>
<td>39.4</td>
<td>35</td>
<td>36.6</td>
</tr>
<tr>
<td>3</td>
<td>36.0</td>
<td>14</td>
<td>36.5</td>
<td>25</td>
<td>36.4</td>
<td>36</td>
<td>38.0</td>
</tr>
<tr>
<td>4</td>
<td>35.7</td>
<td>15</td>
<td>36.7</td>
<td>26</td>
<td>38.1</td>
<td>37</td>
<td>37.3</td>
</tr>
<tr>
<td>5</td>
<td>39.0</td>
<td>16</td>
<td>37.2</td>
<td>27</td>
<td>36.3</td>
<td>38</td>
<td>37.3</td>
</tr>
<tr>
<td>6</td>
<td>36.1</td>
<td>17</td>
<td>36.5</td>
<td>28</td>
<td>39.9</td>
<td>39</td>
<td>38.0</td>
</tr>
</tbody>
</table>

Table–3.1.1

Observed Value on Annual Maximum of Ambient Air Temperature (in Degree Celsius)
Here the observed values $x_i$ ($i = 1, 2, 3, \ldots, 43$) can be assumed to be composed of a parameter $\mu$ (representing the central tendency of annual maximum) and chance errors.

### 3.2 Evaluation of Value of $\mu$ (the central tendency of annual maximum)

The computed values of $A_n = A(x_1, x_2, \ldots, x_n)$ & $G_n = G(x_1, x_2, \ldots, x_n)$ have been shown in Table–3.1.2.

In Table–3.1.2, it is found that the values of $A_n = A(x_1, x_2, \ldots, x_n)$ and $G_n = G(x_1, x_2, \ldots, x_n)$ are approaching 37.2.

Hence, with probability approaching 1, the value of the central tendency of annual maximum of the ambient air temperature at Guwahati can be taken as 37.2 Degree Celsius.

However, 37.2 Degree Celsius cannot be the actual value of the central tendency with probability

$$(\frac{1}{2})^{42} = 2.273736754 \times 10^{-13}$$

<table>
<thead>
<tr>
<th>TPR No ($n$)</th>
<th>Value of $A_n = A(x_1, x_2, \ldots, x_n)$</th>
<th>TPR No ($n$)</th>
<th>Value of $A_n = A(x_1, x_2, \ldots, x_n)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>37.1</td>
<td>23</td>
<td>36.92608695652174</td>
</tr>
<tr>
<td>2</td>
<td>36.85</td>
<td>24</td>
<td>37.02916666666667</td>
</tr>
<tr>
<td>3</td>
<td>36.5666666666667</td>
<td>25</td>
<td>37.004</td>
</tr>
<tr>
<td>4</td>
<td>36.35</td>
<td>26</td>
<td>37.04615384615385</td>
</tr>
<tr>
<td>5</td>
<td>36.88</td>
<td>27</td>
<td>37.01851851851852</td>
</tr>
<tr>
<td>6</td>
<td>36.75</td>
<td>28</td>
<td>37.12142857142857</td>
</tr>
<tr>
<td>7</td>
<td>37.1</td>
<td>29</td>
<td>37.13103448275862</td>
</tr>
<tr>
<td>TPR No (n)</td>
<td>Value of ( G_n = G(x_1, x_2, \ldots, x_n) )</td>
<td>TPR No (n)</td>
<td>Value of ( G_n = G(x_1, x_2, \ldots, x_n) )</td>
</tr>
<tr>
<td>------------</td>
<td>---------------------------------</td>
<td>------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>1</td>
<td>37.10000000000000000000000000000001</td>
<td>23</td>
<td>36.907951375530228273997704899203</td>
</tr>
<tr>
<td>2</td>
<td>36.84915195767397816492457471162</td>
<td>24</td>
<td>37.008568334146486997509657949442</td>
</tr>
<tr>
<td>3</td>
<td>36.563898828796401481877921649392</td>
<td>25</td>
<td>36.984031371155694976981399533444</td>
</tr>
<tr>
<td>4</td>
<td>36.345983740858013867930065117099</td>
<td>26</td>
<td>37.026342595226743557691601271825</td>
</tr>
<tr>
<td>5</td>
<td>36.861929466436019045199560948853</td>
<td>27</td>
<td>36.99918361772609404713638815137</td>
</tr>
<tr>
<td>6</td>
<td>36.73383352490652952733500969237</td>
<td>28</td>
<td>37.099058031246057185229984821213</td>
</tr>
<tr>
<td>7</td>
<td>37.076408237779582297449661223967</td>
<td>29</td>
<td>37.109394916981577631613488308734</td>
</tr>
<tr>
<td>8</td>
<td>37.3115700555521037921393145135076</td>
<td>30</td>
<td>37.122349300390835138945562995519</td>
</tr>
<tr>
<td>9</td>
<td>37.082517567449318065559488651693</td>
<td>31</td>
<td>37.108649562396908895644952409924</td>
</tr>
<tr>
<td>10</td>
<td>37.054168483041978324085190136495</td>
<td>32</td>
<td>37.063799115887290185156645453858</td>
</tr>
</tbody>
</table>
3.3. Annual Minimum of Ambient Air Temperature at Guwahati:

The following table (Table–3.2.1) shows the observed data on annual maximum of ambient air temperature, occurred in temperature periodic year (TPR), at Guwahati during the period from 1969 to 2013.

As earlier, the observed values

\[ x_i \ (i = 1, 2, 3, \ldots, 43) \]

can in this case also be assumed to be composed of a parameter \( \mu \) (representing the central tendency of annual minimum) and chance errors.

**Table–3.2.1**

<table>
<thead>
<tr>
<th>TPR No ((i))</th>
<th>Observed Value ((x_i))</th>
<th>TPR No ((i))</th>
<th>Observed Value ((X_i))</th>
<th>TPR No ((i))</th>
<th>Observed Value ((x_i))</th>
<th>TPR No ((i))</th>
<th>Observed Value ((X_i))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.6</td>
<td>12</td>
<td>7.5</td>
<td>23</td>
<td>5.9</td>
<td>34</td>
<td>8.0</td>
</tr>
<tr>
<td>2</td>
<td>5.9</td>
<td>13</td>
<td>8.3</td>
<td>24</td>
<td>8.4</td>
<td>35</td>
<td>7.9</td>
</tr>
<tr>
<td>3</td>
<td>8.2</td>
<td>14</td>
<td>4.9</td>
<td>25</td>
<td>7.8</td>
<td>36</td>
<td>6.7</td>
</tr>
<tr>
<td>4</td>
<td>5.0</td>
<td>15</td>
<td>6.1</td>
<td>26</td>
<td>7.5</td>
<td>37</td>
<td>9.6</td>
</tr>
</tbody>
</table>
3.4. Determination of Value of $\mu$ (the central tendency of annual minimum)

The computed values of $A_n = A(x_1, x_2, \ldots, x_n)$ and $G_n = G(x_1, x_2, \ldots, x_n)$ have been shown in Table–3.2.2.

In Table–3.2.2, it is found that the values of $A_n = A(x_1, x_2, \ldots, x_n)$ and $G_n = G(x_1, x_2, \ldots, x_n)$ are not approaching a common value.

Thus, either the data are insufficient to yield the true value of the central tendency of annual minimum of the ambient air temperature at Guwahati or the data do not follow the model described by equation (2.1).

<table>
<thead>
<tr>
<th>TPR No (n)</th>
<th>Values of $A_n = A(x_1, x_2, \ldots, x_n)$</th>
<th>Values of $A_n = A(x_1, x_2, \ldots, x_n)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.6</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>6.6</td>
<td>24</td>
</tr>
<tr>
<td>3</td>
<td>6.36666666666666666667</td>
<td>25</td>
</tr>
<tr>
<td>4</td>
<td>6.825</td>
<td>26</td>
</tr>
<tr>
<td>5</td>
<td>6.46</td>
<td>27</td>
</tr>
<tr>
<td>6</td>
<td>6.4333333333333333333333</td>
<td>28</td>
</tr>
<tr>
<td>7</td>
<td>6.571428571428571428571</td>
<td>29</td>
</tr>
<tr>
<td>8</td>
<td>6.575</td>
<td>30</td>
</tr>
<tr>
<td>TPR No (n)</td>
<td>Values of $G_n = G(x_1, x_2, \ldots, x_n)$</td>
<td>TPR No (n)</td>
</tr>
<tr>
<td>-----------</td>
<td>------------------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>1</td>
<td>6.600000000000000000000000000000002</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>6.2401923047290778184540741626669</td>
<td>24</td>
</tr>
<tr>
<td>3</td>
<td>6.8349697977658553153553991473341</td>
<td>25</td>
</tr>
<tr>
<td>4</td>
<td>6.32113331345946749216627042232</td>
<td>26</td>
</tr>
<tr>
<td>5</td>
<td>6.3169010013367452786283625052215</td>
<td>27</td>
</tr>
<tr>
<td>6</td>
<td>6.4857272185701169855041242093667</td>
<td>28</td>
</tr>
<tr>
<td>7</td>
<td>6.5019299609413096599434266552134</td>
<td>29</td>
</tr>
<tr>
<td>8</td>
<td>6.463398937912701904703289281854</td>
<td>30</td>
</tr>
<tr>
<td>9</td>
<td>6.551405846887191769812286015146</td>
<td>31</td>
</tr>
<tr>
<td>10</td>
<td>6.5153871170191827914459649178302</td>
<td>32</td>
</tr>
<tr>
<td>11</td>
<td>6.5048119729138514238168851259175</td>
<td>33</td>
</tr>
<tr>
<td>12</td>
<td>6.582440726934850237656680461569</td>
<td>34</td>
</tr>
<tr>
<td>13</td>
<td>6.7008890789412278622783766593454</td>
<td>35</td>
</tr>
</tbody>
</table>

Table 3.2.3
14 6.5527362805560146596136072863623 36 7.2944086947827372183189849879502
15 6.5215350514733093207287599447551 37 7.3548715998389568402950960158236
16 6.5949101184036978619969386660223 38 7.81546129430738975137961946202
17 6.6987014182595463410514031559311 39 7.7934100682283330885886838578
18 6.7507455267102729305100790423158 40 7.7551256022354259620074788758758
19 6.8616303970395831831869394230887 41 7.6937848209202825772540315655963
20 6.8534570444635167207666214239384 42 6.8994798638029492332170025993497
21 6.927944662370421736229802749321 43 6.9562840436945652525470112265315
22 6.948733241847847197680953313006 44 6.9882108302873798619833480810597

Note: Data corresponding to TPR 29, 30 &31 are not available. Therefore, the values of
\[ A_n = A(x_1, x_2, \ldots, x_n) \] and \[ G_n = G(x_1, x_2, \ldots, x_n) \]
could not be computed corresponding these three TPR.

4. Conclusion:

The method, developed here, can provide the value of the parameter if the data follow the model
described by equation (2.1) and if the data size is sufficiently large for obtaining the common converging
point of \[ A_n = A(x_1, x_2, \ldots, x_n) \] and \[ G_n = G(x_1, x_2, \ldots, x_n) \). Conversely, if the
common converging point of \[ A_n = A(x_1, x_2, \ldots, x_n) \] and \[ G_n = G(x_1, x_2, \ldots, \]
\( x_n \) is not achieved from the set of data then it implies that either the data do not follow the model
described by equation (2.1) or the data size is not sufficient to yield the common converging point.

Regarding the findings obtained on annual maximum and annual minimum of ambient air temperature
at Guwahati, the following conclusion can be drawn:

4.1. The central tendency of Annual Maximum of Ambient Air Temperature at Guwahati can be taken as
37.2 Degree Celsius, with probability approaching 1 (i.e. with probability certainty), since all the
methods applied have yielded the same numerical results and thus the corresponding data can be treated
to follow the model described by equation (2.1).

4.2. The central tendency of Annual Minimum of Ambient Air Temperature at Guwahati is not
determinable since the methods applied have yielded different numerical results and thus the
corresponding data cannot be treated to follow the model described by equation (2.1).
In this connection, it is to be mentioned that the central tendency of each of annual maximum and annual minimum of the ambient air temperature at Guwahati has been evaluated by another method already developed (Chakrabarty, 2017). The findings obtained by that method have been found to be identical with the findings obtained by the method developed here.

References:


