<table>
<thead>
<tr>
<th>Title</th>
<th>Unbiasedness in the test of goodness of fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Okamoto, Masashi</td>
</tr>
<tr>
<td>Citation</td>
<td>Osaka Mathematical Journal. 4(2) P.211-P.214</td>
</tr>
<tr>
<td>Issue Date</td>
<td>1952</td>
</tr>
<tr>
<td>Text Version</td>
<td>publisher</td>
</tr>
<tr>
<td>URL</td>
<td><a href="https://doi.org/10.18910/4183">https://doi.org/10.18910/4183</a></td>
</tr>
<tr>
<td>DOI</td>
<td>10.18910/4183</td>
</tr>
<tr>
<td>rights</td>
<td></td>
</tr>
</tbody>
</table>
Unbiasedness in the Test of Goodness of Fit

By Masashi OKAMOTO

1. Introduction. Let \( X_1, \ldots, X_N \) be a random sample from the population with the d.f. \( F(x) \). We are asked to test the hypothesis \( H_0 \) that \( F(x) \) is identical with a specified continuous d.f. \( F_0(x) \) against all alternatives. For this purpose we shall use the multinomial distribution, dividing the real line into \( n \) intervals \((a_{i-1}, a_i], i = 1, \ldots, n\), where \( a_0 = -\infty \) and \( a_n = +\infty \), so that \( F_0(a_i) - F_0(a_{i-1}) = 1/n, i = 1, \ldots, n \). If \( a_i \) are not determined uniquely, we may take any values satisfying the conditions. Put \( p_i = F(a_i) - F(a_{i-1}) \) and denote by \( N_t \) the number of \( X \)'s that fall into the interval \((a_{i-1}, a_i] \). Then, of course, \( \sum_{i=1}^{n} p_i = 1 \) and \( \sum_{i=1}^{n} N_t = N \). Denote, further, by \( W \) the space consisting of \( n \)-dimensional lattice points \((k_1, \ldots, k_n)\), where \( k_i \) is regarded as the observed value of the random variable \( N_t \) (therefore, \( \sum_{i=1}^{n} k_i = N \)).

The test is equivalent with determining the set (acceptance region) in the space \( W \). The set \( S \) in \( W \) will be called symmetric provided that, if \( S \) contains the point \((k_1, \ldots, k_n)\), then \( S \) contains also all its permutations \((k_1', \ldots, k_n')\). We shall say, finally, that \( S \) satisfies condition \( O \) when, if \( S \) contains \((k_1, \ldots, k_n)\) such as \( k_j \geq k_i + 2 \), then \( S \) contains also \((k_1, \ldots, k_i+1, \ldots, k_j-1, \ldots, k_n)\). It is easily verified that if \( S \) is symmetric the convexity implies the condition \( O \). The converse, however, is not necessarily true. For example, we shall consider, in the case \( N = 12, n = 3 \), the set \( S \) consisting of nine points shown in Fig. 1 and their permutations. \( S \) is symmetric and satisfies the condition \( O \), but is not convex, since the middle point \((7, 4, 1)\) of the points \((8, 2, 2), (6, 6, 0)\) does not belong to \( S \).

2. Theorem of unbiasedness.

Theorem. If the acceptance region \( R \) of the test is symmetric and satisfies the condition \( O \), the test of \( H_0 \) is unbiased against any alternative.

Proof. Putting
we have to prove that \( P \) is maximum when \( p_1 = \ldots = p_n \).

Since \( R \) is symmetric, \( P \) is a symmetric function in \( p_1, \ldots, p_n \).
Thus we have only to prove that, if \( p_1 \leq p_2 \),

\[
P(x) = \sum_{(k_1, \ldots, k_n) \in R} \frac{N!}{k_1! \ldots k_n!} (p_1 + x)^{k_1} (p_2 - x)^{k_2} p_3^{k_3} \ldots p_n^{k_n}
\]
is monotonically increasing in \( x \), when \( 0 \leq x \leq (p_2 - p_1)/2 \). In the sequel we shall consider \( x \) only in this range.

For any \((n-1)\)-tuple \((k, k_3, \ldots, k_n)\) such that \( k + k_3 + \ldots + k_n = N \), let \( R_{kk_3 \ldots k_n} \) be the subset of \( R \) consisting of \((k_1, \ldots, k_n)\) such that \( k_1 + k_2 = k \). (Some may be null set.) Then \( R_{kk_3 \ldots k_n} \) are disjoint and exhaust \( R \). Therefore

\[
P(x) = \sum_{(k_1, \ldots, k_n) \in R} P_{kk_3 \ldots k_n}(x)
\]

where

\[
P_{kk_3 \ldots k_n}(x) = \sum_{k_1, \ldots, k_n} \frac{N!}{k_1! \ldots k_n!} (p_1 + x)^{k_1} (p_2 - x)^{k_2} p_3^{k_3} \ldots p_n^{k_n},
\]

\( \sum \) extending over all \( n \)-tuples \((k_1, \ldots, k_n)\) belonging to \( R_{kk_3 \ldots k_n} \).

Since \( R \) is symmetric and satisfies the condition \( O \), all \( R_{kk_3 \ldots k_n} \) are symmetric and satisfy the condition \( O \) with respect to \( k_1, k_2 \). Thus, if not null set,

\[
R_{kk_3 \ldots k_n} = \left\{ (i, k-i, k_3, \ldots, k_n) : j \leq i \leq k-j \right\},
\]

where \( j \) is a non-negative integer \( \leq k/2 \), depending on \( k, k_3, \ldots, k_n \), and so

\[
P_{kk_3 \ldots k_n}(x) = \sum_{i=j}^{k-j} \frac{N!}{i! (k-i)! k_3! \ldots k_n!} (p_1 + x)^{i} (p_2 - x)^{k-i} p_3^{k_3} \ldots p_n^{k_n}
\]

\[
= \frac{N!}{k! k_3! \ldots k_n!} p_3^{k_3} \ldots p_n^{k_n} \sum_{i=j}^{k-j} \binom{k}{i} (p_1 + x)^{i} (p_2 - x)^{k-i}.
\]

Put

\[
B_i(x) = \sum_{i=j}^{k-j} \binom{k}{i} (p_1 + x)^{i} (p_2 - x)^{k-i}.
\]

If \( j = 0 \),

\[
B_0(x) = \sum_{i=0}^{k-j} \binom{k}{i} (p_1 + x)^{i} (p_2 - x)^{k-i} = (p_1 + p_2)^k.
\]

If \( 1 \leq j \leq k/2 \), denoting by the prime the derivative with respect to \( x \).

\[
B_j(x) = \frac{k!}{(j-1)! (k-j)!} \left( (p_1 + x)^{j-1} (p_2 - x)^{k-j} - (p_1 + x)^{k-j} (p_2 - x)^{j-1} \right).
\]
Unbiasedness in the Test of Goodness of Fit

Since \( j - 1 \leq k - j \), \( p_1 + x \leq p_2 - x \), we have

\[
B_j(x) \geq 0, \quad \text{for} \quad 1 \leq j \leq k/2.
\]

With (2), (3), (4) and (5), we have

\[
P_{k_k \ldots k_n}(x) \geq 0.
\]

This and (1) imply

\[
P'(x) \geq 0,
\]

and this completes the proof.

3. Applications.

(1) The \( \chi^2 \)-test. The acceptance region \( R \) of the \( \chi^2 \)-test consists of the points \( (k_1, \ldots, k_n) \) such that

\[
\sum_{i=1}^{n} (k_i - N/n)^2 \leq c^2,
\]

where \( c \) is a constant depending on the level of significance of the test. \( R \) is readily seen to be symmetric. In order to verify the condition \( O \), we have only to show that

\[
(k_1 + 1 - N/n)^2 + (k_2 - 1 - N/n)^2 \leq (k_1 - N/n)^2 + (k_2 - N/n)^2,
\]

when \( k_2 \geq k_1 + 2 \). This inequality follows from the relations

\[
(k_1 + 1 - a)^2 - (k_1 - a)^2 = 2k_1 + 1 - 2a
\]

\[
\leq 2k_2 - 1 - 2a = (k_2 - a)^2 - (k_2 - 1 - a)^2.
\]

Thus, by the theorem of the preceding section, the \( \chi^2 \)-test of \( H_0 \) is unbiased. This fact was mentioned by H. B. Mann and A. Wald [1], but as they used the Taylor expansion of the power, it is only the local unbiasedness that they proved.

(2) David’s test. The acceptance region \( R \) of David’s test [2] consists of \( (k_1, \ldots, k_n) \) such that at most \( c \) \( k \)'s are zero, where \( c \) is again a constant depending on the level of significance.

\( R \) is symmetric. As for the condition \( O \), let \( A = (k_1, \ldots, k_n) \in R \) and \( k_j \geq k_i + 2 \). If \( k_i = 0 \), the number of zeroes in \( B = (k_1, \ldots, k_i + 1, \ldots, k_j - 1, \ldots, k_n) \) is smaller by one than that in \( A \). If \( k_i > 0 \), both are equal. Therefore \( B \in R \), and the condition \( O \) is satisfied.

Thus, David’s test is also unbiased. The author proved it in his recent paper [3], but the proof was lacking in generality and simpleness.

(Received July 10, 1952)
References

