Calculation of Normal Ranges by Methods Used for Resolution of Overlapping Gaussian Distributions

E. Melvin Gindler

Distribution of a quantity, e.g., concentration of a serum constituent, in a typical general hospital population is considered. It is assumed that the distribution within any subpopulation is Gaussian and that adjacent subpopulations overlap somewhat, presenting overlapping Gaussian distributions. Bhattacharya's procedure for resolving such overlapping distributions, based upon differentiation of the Gaussian distribution equation, is applied to the determination of the apparent normal range as well as, in some cases, an abnormal range. Gaussian probability paper is used for estimating the normal range and the conditions discussed under which this method may be expected to give a valid estimate. Use of the chi-square test to evaluate the long-term constancy of clinical laboratory data distribution, normal and abnormal, is also considered.

Bhattacharya (1) proposed a graphical procedure for the efficient resolving of overlapping Gaussian distributions, a situation frequently encountered in clinical medicine (2, 3). This work represents a solution to a problem considered earlier by several other workers (1), perhaps first by Pearson in 1894 (4). We shall consider the method of Bhattacharya, using the applicable equations in a somewhat modified form and in a way familiar to clinical chemists.

Equations

The Gaussian distribution equation may be used to describe the distribution of a subpopulation, part of a general population. Consider kidney disease as an example, in which the entire patient population is the general population. There is a subpopulation of patients, generally the majority, who are free from kidney disease. These patients would be mainly characterized by their normal serum urea nitrogen concentration. In addition, there would be one or more subpopulations of patients whose serum urea nitrogen concentrations tend to lie outside the normal range, with correspondingly different degrees of kidney damage among some of these subpopulations.

First consider the distribution of \( x \), the laboratory value in the \( m \)th subpopulation, where \( y_m \) is the probability density of the \( m \)th subpopulation

\[
y_m = \frac{A_m}{\sigma_m \sqrt{2\pi}} e^{-\frac{(x - \mu_m)^2}{2\sigma_m^2}}
\]

(1)

and the other terms in the Gaussian distribution equation are \( \mu_m \), the mean laboratory value of the \( m \)th subpopulation; \( \sigma_m \), the standard deviation of the \( m \)th subpopulation; and \( A_m \), the ratio of number of patients in the \( m \)th subpopulation to that in the total population.

The observed probability density function, \( y \), was defined by Pearson (4) as

\[
y = \Sigma y_m
\]

(2)

For practical laboratory tests it is a necessary condition, when normal and abnormal subpopulations overlap, that this overlapping be only partial. A test in which normal and abnormal data overlap completely is generally useless. Therefore, we may expect that there will be a region for the normal subpopulation, as well as for one or more of the other subpopulations, in which there is virtually no interference from any of the nearby subpopulations. For each subpopulation this clear region is generally within the range \( x = \mu_m \pm 2 \sigma_m \). All of the other subpopulations have probability density

---

From the Department of Pathology, Rockford Memorial Hospital, Rockford, Ill. 61103.
Abstracted in part in Chicago Clinical Chemist 1, 6 (1968).
Received Oct. 2, 1969; accepted Dec. 2, 1969.
values that are very small. Then, if the second subpopulation, \( m = 2 \), is the normal population and \((\mu_2 + 2 \sigma_2) > x > (\mu_2 - 2 \sigma_2)\), then \( y \cong y_2 \).

The substitution \( y = y_m = y_2 \) may be made in Equation 1. This equation may then be written using the natural logarithm of \( y \), \( \ln y \), and differentiated with respect to \( x \). Since the equations are to be used for finite increments rather than infinitesimal increments, the substitution \( \Delta \ln y/\Delta x \) is used instead of \( d \ln y/dx \). Uniform \( \Delta x \) intervals of size \( h \) are used. The final equations of Bhattacharya that result are

\[
\mu_2 = \lambda_2 + h/2 \tag{3}
\]

and

\[
\sigma_2^2 = h \Delta x / \Delta (\Delta \ln y) - h^2/12 \tag{4}
\]

The term \( \lambda_2 \) is the value of \( x \) at which \( \Delta \ln y/\Delta x = 0 \). The term \( h^2/12 \) is Sheppard's correction for the grouping (δ).

**Applications of Bhattacharya's Method**

A common situation to be considered here is one in which there are three partially overlapping distributions. For the subpopulation in which the \( x \) values are generally below the normal range we shall set \( m = 1 \). For the normal subpopulation we shall set \( m = 2 \). For the subpopulation in which the \( x \) values are generally above the normal range we shall set \( m = 3 \). We now consider two common procedures, both fitting these conditions: serum urea nitrogen and serum chloride. With the first determination, \( A_2 > A_3 \gg A_1 \) and, the second, \( A_2 > A_1 \cong A_3 \).

---

**Table 1. Urea Nitrogen Concentration Values, Evaluated by the Bhattacharya Method**

<table>
<thead>
<tr>
<th>Urea N interval range, mg/100 ml</th>
<th>( z = ) midpoint of interval range</th>
<th>( y = ) number of patients in interval range</th>
<th>( \Delta \ln y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-4</td>
<td>2</td>
<td>5</td>
<td>1.280</td>
</tr>
<tr>
<td>5-8</td>
<td>6</td>
<td>18</td>
<td>1.312</td>
</tr>
<tr>
<td>9-12</td>
<td>10</td>
<td>67</td>
<td>0.582</td>
</tr>
<tr>
<td>13-16</td>
<td>14</td>
<td>120</td>
<td>-0.245</td>
</tr>
<tr>
<td>17-20</td>
<td>18</td>
<td>94</td>
<td>-0.960</td>
</tr>
<tr>
<td>21-24</td>
<td>22</td>
<td>36</td>
<td>-0.641</td>
</tr>
<tr>
<td>25-28</td>
<td>26</td>
<td>19</td>
<td>-0.639</td>
</tr>
<tr>
<td>29-32</td>
<td>30</td>
<td>10</td>
<td>0.0954</td>
</tr>
<tr>
<td>33-36</td>
<td>34</td>
<td>11</td>
<td>-0.201</td>
</tr>
<tr>
<td>37-40</td>
<td>38</td>
<td>9</td>
<td>-0.405</td>
</tr>
<tr>
<td>41-44</td>
<td>42</td>
<td>6</td>
<td>-0.692</td>
</tr>
<tr>
<td>45-48</td>
<td>46</td>
<td>3</td>
<td>0.845</td>
</tr>
<tr>
<td>49-52</td>
<td>50</td>
<td>7</td>
<td>-0.405</td>
</tr>
<tr>
<td>53-56</td>
<td>54</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Over 56</td>
<td>31</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The \( x \) and \( \Delta \ln y \) values of this table are plotted in Fig. 1.

---

**Results with urea nitrogen data.** These were obtained with a Technicon AutoAnalyzer and the reaction with diacetyl monoxime. Data from 439 patients are listed in Table 1, and the graph derived from them is shown in Fig. 1. The chosen interval, \( h \), was 4 mg/100 ml with a range, \( x \), of 4 to 56 mg/100 ml, sufficient to include at least the entire normal range.

Equation 3 is applied to the graph of Fig. 1, where it is seen that at \( \Delta \ln y/\Delta x = 0 \) (or \( \Delta \ln y = 0 \), since \( \Delta x = h \) on this graph), \( x = \lambda_2 = 13.0 \) mg/100 ml. Then \( \mu_2 = 13.0 + \frac{h}{2} = 15.0 \) mg/100 ml. Equation 4 is applied to Fig. 1 by first evaluating the reciprocal of the slope by finding the ratio of the length of abscissa traversed by the plotted line to the corresponding length of ordinate traversed. This ratio is then directly substituted into Equation 4 along with the value of \( h \). Then \( \sigma_2^2 = 20.0 \) mg/100 ml and \( \sigma_2 = 4.47 \) mg/100 ml, so that \( 2 \sigma_2 = 8.94 \) mg/100 ml \( \cong 9 \) mg/100 ml and the apparent normal range is \( \mu_2 \pm 2 \sigma_2 = 15 \pm 9 \) mg/100 ml. The term, “apparent normal range,” is used because it is uncertain at this point how much interference there is from overlapping nearby abnormal subpopulations.

Inspection of Fig. 1 shows that it is also possible.
to calculate $\mu_3$ and $\sigma_3$, since a second set of points lies along a straight line. The value for $\lambda_3$ is 31.4 mg/100 ml and the foregoing calculation procedure gives $\lambda_3 = 33.4$ mg/100 ml and $\sigma_3 = 7.7$ mg/100 ml. Thus the abnormal range is 18 to 49 mg/100 ml.

The apparent normal range calculated here compares favorably with both the usual normal range of 8 to 25 mg/100 ml, established for more than a decade, and the normal range of 5 to 25 mg/100 ml given by Henry (6).

The proposed procedure was also applied to 4071 consecutive values for urea nitrogen concentration from the University of Wisconsin Hospitals. These values were grouped in numerical order by computer. The normal range was calculated to be 4 to 24 mg/100 ml; the official range at the source laboratory is 3 to 27 mg/100 ml. This close agreement with the results from our patients would be expected; the two hospitals are within 100 miles of each other and serve much the same population.

The abnormal range obtained from our data is presently of unknown importance. It definitely appears that several of these concentration ranges overlap, an observation supported by the data from the University of Wisconsin. Perhaps the only present value of determining the abnormal range is that concentrations in the upper part of the apparent normal range may indicate early kidney disease. We have confirmed this recently (Gindler and Hanson, unpublished work) by determining the serum creatinine for a large number of patients with urea nitrogen concentrations in the upper normal range; a significant number of these patients had abnormally high serum creatinine concentrations. Consequently, we now routinely determine serum creatinine concentration whenever serum urea nitrogen concentration equals or exceeds 22 mg/100 ml.

Serum glucose concentrations overlap considerably between the upper apparent normal range and the adjacent abnormal range. Consequently, it is frequently necessary to differentiate between normal and diabetic patients with more sensitive tests, such as glucose tolerance. This is another example of the fact that a single test cannot be used to determine whether a value in the overlapping area is normal. Although some attempts have been made (2) to differentiate between normal and abnormal values in an overlapping area, there does not seem to be a generally accepted method for doing so.

Results with serum chloride data. These were obtained by mercurimetric titration; the SQBA indicator (7) was used. Figure 2 shows a Bhattacharya graph for 255 consecutive chloride values obtained routinely in our laboratory in November 1968. The calculated normal range is 97 to 107 mEq/liter of serum. Our established normal range is 99 to 105 mEq/liter; Henry (6) gives a normal range of 98 to 109 mEq/liter, in good agreement. The data for 80 of the 255 patients fell below the calculated (apparent) normal range, while those for 42 patients fell above it. Serum chloride concentration was seldom determined except when electrolyte imbalance was suspected, but serum urea nitrogen concentration is determined for all entering patients.

Graphs on Gaussian Probability Paper

Bhattacharya has referred to the use of graphs on Gaussian probability paper in attempting to resolve overlapping Gaussian distributions. Hoffmann (8) independently found a method for determining the normal range by use of a graph on Gaussian probability paper. Graphically, Hoffmann's procedure gives a reliable estimate of the normal range when one of the abnormal subpopulations, $A_1$ or $A_2$, is very small but not when both $A_1$ and $A_2$ are of substantial magnitude. The question of overlapping distributions need not be considered here. On the other hand, it was shown above that the method of Bhattacharya gives reliable estimates of the normal range regardless of the relative values of $A_1$, $A_2$, and $A_3$.

The serum chloride concentration data used for Fig. 2 were also used for Fig. 3, a graph made on Gaussian probability paper. Figure 3 shows an apparent normal range for chloride of 85 to 112 mEq/liter, far broader than the accepted normal ranges for this ion in serum. The suggested (8) $2 \sigma$ width for the normal range was used.

The data of Table 1 were used to prepare the urea nitrogen graph of Fig. 3 and here the apparent normal range $(\pm 2 \sigma)$ was 6 to 26 mg/100 ml.
This calculation, made by the Hoffmann method, agrees closely with the apparent 6 to 24 mg/100 ml normal urea nitrogen range calculated by the Bhattacharya method.

Figure 4 is an attempt to show by graph why the Hoffmann procedure works well with urea nitrogen but poorly with chloride. We consider again a total population having a normal subpopulation in which \( m = 2 \); an abnormal subpopulation in which \( m = 1 \), characterized by abnormally low values; and a second abnormal subpopulation in which \( m = 3 \), characterized by abnormally high values. We shall set \( \sigma = \sigma_1 = \sigma_2 = \sigma_3; \mu_2 - \mu_1 = 2 \sigma \) and \( \mu_2 - \mu_3 = 2 \sigma \). Line A of Fig. 4 is a graph of the situation in which \( A_2 = 1, A_1 = A_3 = 0 \). As expected, line A is straight. Line B of Fig. 4 is a graph of the situation in which \( A_1 = 0, A_2 = 0.9, \) and \( A_3 = 0.1 \). At small values of \( x \), line B is virtually superimposed upon line A, so that extension of the straight part of line B would be expected to give a valid estimate of the apparent normal range. The situation in which \( A_1 \) is a very small value is analogous to that for serum urea nitrogen, so the Hoffmann procedure would be expected to give a valid estimate of the normal range.

Line C of Fig. 4 represents a situation in which \( A_1 = A_2 = 0.05 \) and \( A_3 = 0.9 \). Line C is essentially S-shaped and is hardly straight anywhere, except near its inflection point. Line C is also superimposed upon the midsection of line A for a short distance. (Here line A represents the correct line for the estimation of the apparent normal range throughout its length.) Hence, if neither \( A_1 \) nor \( A_3 \) is infinitesimally small, as in the case for serum chloride concentration, it would be expected that the Hoffmann procedure would give an erroneously broad estimate of the apparent normal range.

Figure 4 thus indicates that the Hoffmann procedure must be used with caution. However, once it is established that \( A_1 \) (in urea nitrogen) or \( A_3 \) (in serum albumin) is sufficiently small for a given test procedure, then the Hoffmann technique will doubtless continue to be used for such a procedure with good results, because it is considerably less laborious than the Bhattacharya procedure. The Bhattacharya procedure will probably be used more with complicated distributions among several subpopulations.

**Application of Chi-Square Test**

Periodic evaluation of the normal range over a number of years may serve as an index of laboratory precision if it is certain that the hospital population is essentially unchanging. A drawback to using only the normal value ranges is that they necessarily traverse only a small part of the range actually observed in the laboratory. A technique which permits the evaluation of the constancy of a distribution, by seeing how well it agrees with some standard distribution, is the well-known chi-square test \( (9) \) defined by the equation

\[
\chi^2 = \sum [(W - E)^2/E]
\]

where \( W \) is the observed number of patients in a
given interval of test values, and $E$ is the expected number of patients in the same interval of test values.

$E$ values were found for serum urea nitrogen values obtained nearly two decades ago. Then $W$ values for a few years (up to 1968) were substituted into Equation 5. On comparing the $\chi^2$ values found with those listed in standard tables, it was seen that there has been no significant change in the distribution of our serum urea nitrogen concentrations. These calculations were made from unselected data. Serum glucose concentrations may be changed by emotional state, diet, and medication, and the chi-square test showed that there usually were significant differences in distribution from one day to another.

The author thanks M. A. Evenson, University of Wisconsin Hospital, Madison, for the urea nitrogen data from that institution. Supported in part by Grant H-17, Health Research and Services Foundation, Pittsburgh, Pa.

References