of this book he treats *inter alia* Boltzmann's H-theorem by means of the grand canonical ensemble. In doing so, he generalizes the η and ψ of his canonical ensemble to H and Ψ, where the mean value of H is Boltzmann's function. He shows on p. 200 that these quantities correspond to the negative entropy and the free energy of thermodynamics. As seen from the way it is introduced, the symbol H is clearly meant as a capital η. This is also definitely verified by the fact that this H, like all other capital Greek letters in the book, is printed vertical (nonslanted), while capital Latin letters are printed with italics (slanted types), as illustrated by the symbols on p. 200, and by the Latin E on p. 158 compared to the Greek E (epsilon) on p. 177. We may also note that Zermelo in his German translation of the book in 1905 also maintains the same classification of the capital letters, although by other typographical means.

The given graphical evidence, of which a detailed account is presented elsewhere,[6] seems to leave no reasonable doubt that during the decade before Boltzmann's death in 1906 at least he himself, Gibbs, and Zermelo meant a capital eta when they wrote H for Boltzmann's function. This suggests that the whole problem of the adoption of the symbol H has a connection with Gibbs and his ensemble-statistical ideas, surely well known to leading scientists long before the publication in 1902.

[1]L. Boltzmann, Wien Ber. **66**, 275 (1872).

[2]L. Boltzmann, *Vorlesungen über Gastheorie, I Theil* (Barth, Leipzig, 1896).

[3]S. G. Brush, Arch. Hist. Exact Sci. **12**, 1 (1974).

[4]S. G. Brush, Am. J. Phys. **35**, 892 (1967).

[5]J. W. Gibbs, *Elementary Principles in Statistical Mechanics* (Scribner, New York, 1902); reprinted in facsimile in *The Collected Works of J. Willard Gibbs, Vol. II* (Yale U. P., New Haven, CT, 1928).

[6]S. Hjalmars, TRITA-MEK-76-01, Technical Reports from the Royal Institute of Technology, Department of Mechanics, S-10044 Stockholm, Sweden. (Free of cost on request from the Department.)

When a Gaussian distribution won't do: A short comment on statistics

D. W. Schlitt

Department of Physics and Astronomy, University of Nebraska–Lincoln, Lincoln, Nebraska 68588

(Received 30 January 1976; revised 15 April 1976)

The normal or Gaussian probability distribution rests at the foundation of most of the statistical methods familiar to physicists. The χ^2 test, the F test, Student's t test, etc., each rest on the assumption that some specific quantity has a normal distribution.

What do we do if we know that this assumption of a normal distribution is not valid. A recent paper in this Journal[1] spoke to this question for the case of a Poisson distribution. The author represented the deviation of the Poisson distribution from the normal distribution in terms of a polynomial and made some interesting comments on his result. He did not, however, apply his result to any practical problem in statistics. This is probably because the result is not convenient to use for such things as estimating lifetimes.[2] The purpose of this note is to point out some statistical methods which are convenient to use and to give a simple illustration of the use of one of them.

Statisticians are well aware that not all populations are normally distributed. They have proceeded to develop techniques to use in such cases. Because they often are not discussed in introductory courses in statistics, they have only slowly come into common use in physics.

In order to avoid the assumption of a normal distribution or any specific distribution, a group of methods known as distribution-free or nonparametric methods have been developed; the idea of the robustness of a statistical test has been introduced. A test is robust if it is independent or insensitive to departures from the assumed distribution. A discussion of such tests (directed at physicists) can be found in Ref. 3 and more detailed discussion is given in Ref. 4. The χ^2 test can be replaced by such distribution-free tests as the Cramer–von Mises test or the Kolmogorov–Smirnov test. An example of a recent application of these methods in physics can be found in the work of Ludlam and Slansky.[5]

A typical physics application of statistics is the extraction of an estimate for a physical quantity from experimental data. An example mentioned in Ref. 1 is the determination of a decay constant from counting data. This is an example of the type we are concerned with. We know the number of counts has a Poisson distribution and not a normal distribution. Use of a least-squares or χ^2 method to find the decay constant would immediately suggest itself. As we will see below, these methods are based on an assumption of a normal distribution.

An alternative which can be used if the probability distribution is known, even if it is not normal, is the maximum likelihood method. It is instructive to look at this method in the case of a decay problem. The calculation is simplified version of example 30.20 of Ref. 6. For those interested in a deeper discussion of this problem an early paper by Peierls[7] is of interest. A recent discussion can be found in Ref. 8.

The experiment we consider is one where we count the number of decays in a sample for a time interval Δt_i centered at a time t_i and obtain n_i counts. We do this for N different intervals. The probability of getting n_i counts in the ith interval has a Poisson probability distribution

$$P_i = (n_i!)^{-1}\mu_i^{n_i} \exp(-\mu_i), \qquad (1)$$

where μ_i is the expected number of counts. If the decay constant is λ, then the expected number of counts is approximately (assuming no background counts in n_i)

$$\mu_i = N_0\lambda\Delta t_i \exp(-\lambda t_i). \qquad (2)$$

We choose this form rather than more exact forms for algebraic simplicity.[9]

The problem is to determine "best" values N_0 and λ from the experimental data. For this we use the principle of maximum likelihood. The probability (or likelihood) of getting a set of counts $\{n_i\}$ is

$$L = \prod_{i=1}^{N} P_i = \prod_{i=1}^{N} (n_i!)^{-1}\mu_i^{n_i}\exp(-\mu_i). \quad (3)$$

We want to choose N_0 and λ so this is a maximum for the measured values of $\{n_i\}$. It is more convenient to do the equivalent problem of maximizing $\ln L$:

$$\ln L = \sum_{i=1}^{N} [n_i \ln\mu_i - \mu_i - \ln(n_i!)]$$
$$= \sum_{i=1}^{N} [n_i \ln N_0 + n_i \ln(\lambda\Delta t_i) - n_i\lambda t_i - N_0\lambda\Delta t_i \exp(-\lambda t_i) - \ln(n_i!)]. \quad (4)$$

The maximum is found by setting the partial derivatives of this with respect to λ and N_0 equal to zero. The resulting equations can be written

$$\sum_{i=1}^{N} [n_i - N_0\lambda\Delta t_i \exp(-\lambda t_i)] = 0 \quad (5)$$

and

$$\sum_{i=1}^{N} [n_i t_i - N_0\lambda t_i\Delta t_i \exp(-\lambda t_i)] = 0. \quad (6)$$

Notice that the first of these equations requires that the total number of counts be equal to the expected total number of counts. The second states that the observed mean decay time is equal to the expected mean decay time. These are the requirements which we would naturally write down. That they are also the result of maximizing the likelihood function is reassuring.

Suppose that instead of the Poisson distribution (1) we had taken a normal distribution

$$P_i = (2\pi)^{-1/2}\sigma_i^{-1}\exp[-(n_i - \mu_i)^2/2\sigma_i^2], \quad (7)$$

where the σ_i are assumed to be known. The equation which replaces (4) is

$$\ln L = -\frac{N}{2}\ln(2\pi) - \sum_{i=1}^{N}\left[\ln\sigma_i + \frac{(n_i - \mu_i)^2}{2\sigma_i^2}\right]$$
$$= \text{const} - \frac{1}{2}\sum_{i=1}^{N}\frac{[n_i - N_0\lambda\Delta t_i \exp(-\lambda t_i)]^2}{\sigma_i^2}. \quad (8)$$

Because we have assumed the σ_i are known, the equation is essentially the same as typical least-squares or χ^2 method. The values to be used for the σ_i are those estimated from the data,[10] which in this case would be $n_i^{1/2}$.

From (8) the equations which determine N_0 and λ are

$$\sum_{i=1}^{N}\frac{\Delta t_i \exp(-\lambda t_i)}{\sigma_i^2}[n_i - N_0\lambda\Delta t_i \exp(-\lambda t_i)] = 0 \quad (9)$$

and

$$\sum_{i=1}^{N}\frac{\Delta t_i \exp(-\lambda t_i)}{\sigma_i^2}[n_i t_i - N_0\lambda t_i\Delta t_i \exp(-\lambda t_i)] = 0. \quad (10)$$

These are not the correct equations to use for our problem, but they are the ones which would result from a least-squares estimate for λ.

Through comparison of this result with the one where a Poisson distribution is assumed [Eqs. (5) and (6)], we see they are the same except for the factor $\Delta t_i\sigma_i^{-2}\exp(-\lambda t_i)$. If this quantity turned out to be independent of i, the resulting λ would be the same. This factor can be estimated. If the distribution is Poisson, the $\sigma_i^2 = \mu_i$ and

$$\Delta t_i\sigma_i^{-2}\exp(-\lambda t_i) = 1/N_0\lambda, \quad (11)$$

which is independent of i. This indicates that little difference between the two methods is expected.

It should be noted that using $\sigma_i^2 = \mu_i$ for this estimate is different from doing the same thing in Eq. (8). Also, it is just an estimate; when using Eqs. (9) and (10) to find λ, we would use $\sigma_i^2 = n_i$ as previously noted.

An alternate estimate throws some light on the i dependence of the factor. Instead of using the expected value of n_i for the estimate, as we did above, we can use the most probable value. The most probable value of n_i for Poisson distribution is $\mu_i - \frac{1}{2}$. This results in an estimate

$$\Delta t_i\sigma_i^{-2}\exp(-\lambda t_i) = [N_0\lambda(1 - 2/\mu_i)]^{-1}. \quad (12)$$

The result depends on i and is very close to the previous estimate when μ_i is large. On the other hand, it illustrates why the least-squares method is less desirable than a maximum likelihood calculation based on the Poisson distribution. The estimate in Eq. (12) indicates that intervals with fewer counts are weighted more heavily than those with large numbers of counts. Nevertheless, in a well designed experiment with a reasonable number of counts in each interval the difference between the two methods will be small.

The observant reader will be surprised by this result. The Poisson distribution is always skewed below the mean while the normal distribution is symmetric. The systematic bias should make itself felt and cause a distinct difference in the estimates of the decay constant.

This qualitative conclusion is wrong because of a very important property of maximum likelihood estimates; remember from what we saw above that this includes least-squares methods. The estimates we obtain are consistent, meaning that the estimate converges toward the true value as the number of observations increases. However, the estimates are not *unbiased.*

If we were to repeat our experiment and get a new estimate of λ, it would be expected to differ from the first value because of the finite size of the "sample." Thus there is a distribution for the values of λ and we can conceive of an expected or average value of λ from this distribution. The difference between this expected value for a given size sample and the true value for the whole population is what we mean by the bias of an estimate. If the estimate is unbiased, then the expected value of the estimate and the true value are the same.

Maximum likelihood estimates are consistent but, in general, biased. The bias in the estimate is the reason that the weighted least-squares method and the Poisson distribution can give nearly the same estimate for λ even though one distribution is skewed and the other symmetric. The bias of the estimates will not necessarily compensate as they do in this case. It is possible to correct for the bias in estimates,[3,6] but it usually is not worth the large effort required.

[1]L. J. Curtis, Am. J. Phys. **43**, 1101 (1975).

[2]It is, however, a useful mathematical relation as is pointed out by H. A.

Gersch, Am. J. Phys. **44**, 885 (1976).
[3]W. T. Eadie, D. Drijard, F. E. James, M. Roos, and B. Sadoulet, *Statistical Methods in Experimental Physics* (North-Holland, Amsterdam, 1971).
[4]J. D. Gibbons, *Nonparametric Statistical Inference* (McGraw-Hill, New York, 1971); M. Hollander and D. A. Wolf, *Nonparametric Statistical Methods* (Wiley, New York, 1973).
[5]T. Ludlam and R. Slansky, Phys. Rev. D **8**, 1408 (1973).
[6]S. L. Meyer, *Data Analysis for Scientists and Engineers* (Wiley, New York, 1975).
[7]R. Peierls, Proc. R. Soc. London A **149**, 467 (1935).
[8]A. H. Jaffey, Nucl. Instrum. Methods **81**, 155 (1970); **81**, 253 (1970).
[9]The exact expression is

$$\mu_i = N_0 \exp[-\lambda(t_i - \Delta t_i/2)] - N_0 \exp[-\lambda(t_i + \Delta t_i/2)].$$

[10]Another approach would be to take $\sigma_i^2 = \mu_i$. The dependence of μ_i on λ and N_0 would then need to be taken into account in finding the maximum. The resulting equations are more complicated than those discussed here.

Term projects: An alternative to term papers in introductory courses

F. R. Hickey
Physics Department, Hartwick College, Oneonta, New York 13820
(Received 22 June 1976; revised 10 September 1976)

Two recent notes[1,2] describe the use of term papers and/or experimental projects in introductory physics courses. For the past four years, I have required a term project in Physics 100, a course in twentieth century physics for non-science majors. Since these projects are quite different from those already described, my experience may be of interest to teachers of similar courses.

The purpose of requiring a project is to help students relate the ideas encountered in the course to their own interests, hobbies, or professional goals. At the beginning of the term, a written list of suggestions for possible projects is given to the students. A term paper, if the topic is unique and original, is an acceptable way to fulfill the requirement, but students are strongly encouraged to explore their creativity and devise a project which holds special interest for them. The viewpoint of the liberal arts major is quite different from that of the physicist and this has resulted in some projects that seemed to me unique and impressive.

Table I. A selection of term projects conceived and executed by students in Physics 100.

Three original piano compositions which used only certain parts of the keyboard corresponding to allowed transitions in the Bohr atom.
An embroidery of formulas and pictures centering around the Sophist quotation "Man is the measure of all things."
An 8 mm film dealing with conservation of energy and momentum.
Three original folk songs, each dealing humorously with a topic from 20th century physics.
A full size comic section of a newspaper. Each of the six "comic strips" treats a different topic covered in the course.
Several original sculptures. The best, made from a solid piece of aluminum, is an interpretation of the Dirac "hole theory."
An orginal story using the concept of time in a manner similar to the way in which L. Durell uses it in *Justine*.
Several Styrofoam models dealing with topics from special relativity.
Scale drawings of a length-contracted Starship Enterprise traveling at speeds corresponding to $\beta = 0$, 0.5, and 0.9.
An original illustrated story starring a well-known cartoon character flying his doghouse at relativistic speeds and encountering the effects of length contraction, time dilation, and velocity addition.

Table I lists some of the best projects. One example of the way in which a student related physics to her discipline is given by the first entry in the table. Ann Cuzzi, a music education major, projected a drawing of the Bohr model of the atom on a piano keyboard. With the nucleus centered at middle C on the keyboard, only several notes on each side of middle C are included in the region between any given pair of allowed orbits. Using only those piano keys which lie in the regions corresponding to the energy change related to the first three wavelengths of the Lyman, Balmer, and Paschen series, respectively, she composed three original musical pieces. Other projects listed in the table present a sampling of ways in which other students have related the course material to their hobbies or personal interests.

To insure that students begin to work on their project early in the term, a brief outline is due on a specific date approximately five weeks after the beginning of the term. This also serves to establish communication between me and each student as to the acceptability and judiciousness of the topic chosen. Once a topic has been mutually agreed upon, it cannot be changed without permission.

The most difficult task for the instructor is evaluating the projects. Grading is necessarily subjective and is based on the student's originality, creativity and the degree to which he/she accomplishes the goal previously agreed upon. The projects count as the equivalent of one written examination or approximately 25% of the final grade. Our college has a provision for adding a written comment to a student's transcript in addition to the grade, and I make it known to the class that I will write a comment for outstanding projects.

In general, students enjoy working on the projects, and sometimes present them to the class during the last week of the term. In the words of one student: "Being and feeling rewarded for creativity was a refreshing and beautiful thing."

[1]W. S. Williams, Am. J. Phys. **43**, 550 (1975).
[2]R. A. Brown, Am. J. Phys. **44**, 393 (1976).