PORTING A PEAK SEARCH PROGRAM TO A PC,
NON-LINEAR FIT TESTING

G. W. HAVERLAND

University of Alberta SLOWPOKE Facility, Edmonton, AB (Canada)

(Received December 21, 1993)

The University of Alberta SLOWPOKE Facility has a need for an automated peak searching program. This need was solved by porting and modifying the source code to the Nuclear Data Basic Physics package program. This paper deals specifically with the testing of the non-linear fitting part of the program, which was extensively modified.

A number of years ago the University of Alberta SLOWPOKE Facility purchased the source code to the Nuclear Data Basic Physics package (ND) to carry out peak searching offline. This source code was (easily) ported to a PDP-11 running RT-11, and with a little more difficulty ported to the University of Alberta's Amdahl mainframe computer. The source code was left in FORTRAN (largely FORTRAN IV) with only slight changes to the input and output code.

The code was also ported to MS-DOS. The program input/output was changed and the program was made more modular. This modularity was achieved in a largely automated way. Firstly the source code was run through the UNIX program struct to produce a structured version of the source code in RATFOR (RATIONAL FORTRAN), and then this RATFOR source was converted to FORTRAN-77 by a customised RATFOR preprocessor which was written by the author. This version was not used much, due to the difficulty in transferring spectra from the spectrometers to a PC. It was noted during this port by the author that documentation of the program was severely lacking.

With falling budgets, automation of this system was more desirable. This was accomplished by porting the ND code to the QNX operating system, with the program source code being in ANSI C instead of FORTRAN. Many changes were made to the code with this port. Runtime memory allocation for storing the counts per channel information (and other variables) was one major change. Parsing the program command line for program switches and arguments, such as to control output verbosity, to decouple peak or background shape from the program structure, and to control program operation were also made. Finally some changes were made to use better algorithms. Most of the latter changes were made to the non-linear fitting part of the program, which is the object of this discussion.

Experimental

Almost any text on gamma spectroscopy presents information on what the ideal gamma photopeak should look like, and some mention modifications to real photopeaks [1]. The ND routines assumed a Gaussian peak on a linear (straight line) background and, if the peak was fit, it was fit using some (undocumented) version of a least squares fitting process. Since the peak centroid and

1Trademark, Canberra-Packard Ltd.
2Trademark, Digital Equipment Corporation
3Trademark, Amdahl Corporation
4Trademark, Unix System Labs
5Trademark, Quantum Software Systems

Elsevier Science S. A., Lausanne
Akadémiai Kiadó, Budapest
characteristic width are nonlinear functions of the counts observed in any channel of the region of interest, an iterative process is necessary to fit a peak to a region of interest. The formula for the counts, \(f(z_i)\) in channel \(z_i\) for a Gaussian peak is shown below:

\[
f(z_i) = \frac{A}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z_i - \langle x \rangle)^2}{2\sigma^2}}
\]

where \(A\) is the area of the peak, \(\sigma\) is the characteristic width of the peak (FWHM \(\approx 2.35 \sigma\)), and \(\langle x \rangle\) is the centroid of the peak. It was decided to change the non-linear fitting routine to a Levenberg-Marquardt method (L-M) [2], or a modification of the L-M algorithm.

The ported program peak shape was left as Gaussian, and the background left as a straight line. However, modules that know the functional form of a peak and its background are segregated and written for (hopefully) easy modification if the need arises.

Some of the ported program source code was tested in a methodical manner, as outlined in [3], but more extensive testing was done using Monte Carlo techniques to determine the confidence of fit. The non-linear fit routine (basically as written in [2] except translated to ANSI C and lightly modified) works quite well when the calculated Gaussian peak is fit, even when the initial estimated parameters are significantly different from best estimates. However gamma spectra acquired with multi-channel analysers are discrete entities. The number of counts observed is a positive integer and these data are placed in discrete bins. This discreteness was simulated by generating synthetic spectra by using a uniform (0,1) random number generator to produce Poisson deviates by the rejection method.

The test routine fit 1000 peaks with an expected area of e^2, then 1000 peaks with an expected area of e^{2.5}, ..., e^{19.5}, all on a background of zero counts. It was felt that this range of peak sizes would certainly encompass the range of peak sizes seen in the laboratory. Working with non-zero backgrounds was not done, as the non-linear fit package is designed to work with net counts, and the routine does make use of the estimated variance in each channel during the fitting process.

Results

The degree of confidence recognised by the Monte Carlo experiments is a function of the random deviates generated. It is known that a Poisson distribution asymptotically approaches the Gaussian distribution [4]. Therefore it is expected that a good test for the Poisson number generator would be to test for the following conditions: 1) that an observed distribution for a small mean is the same as theoretically expected, and 2) the variance of a distribution for a large mean is equal to the mean. Another requirement is that the uniform number generator that drives the Poisson generator does not have serial correlation. For a small mean, the observed histogram was compared to the theoretically expected distribution using a Kolomogorov-Smirnov (K-S) test. For a large mean, a test was made to compare the variance with the expected mean using an F test. No test was made for serial correlation in the uniform random number generator (ranl from [2]), since it is advertised as being a very good one, with a very long period and essentially no serial correlation, and is machine independent.

The small mean test was made on the sum of count data in a 21 channel wide region of interest with a characteristic width of 1 channel and an expected sum of counts of e^2. A K-S statistic for the mean of e^2 is 0.017422. The critical value of the K-S statistic for 17 data points at the 90% confidence level is 0.28627 [5], which is much larger than the value observed. The probability of obtaining a K-S distance larger than what was observed was calculated to be \(\approx 1\). The observed and expected sum of count distributions are shown in Figure 1. Another test is to compute the \(\chi^2\) statistic, which is shown below:

\[
\chi^2 = \sum_i \frac{(N_i - n_i)^2}{n_i}
\]

where \(N_i\) is the observed number of events per bin, and \(n_i\) is the expected number of events per bin. For the mean=e^2 data, the \(\chi^2\) statistic is \(\approx 10.19\). The critical value of the \(\chi^2_{0.05}(16)\) is 10.08, which is approximately equal to the calculated value.