Estimating Uncertainty in Physical Measurements, Observational and Environmental Studies: Lessons From Trends in Nuclear Data

Alexander Shlyakhter[1], Ilya Shlyakhter[2], Claire Broido[1], and Richard Wilson[1]

[1]Department of Physics
and
Northeast Regional Center for Global Environmental Change
Harvard University, Cambridge, MA 02138 USA

[2]Brookline High School, Brookline, MA 02146 USA

Abstract

The time trends in the sequential series of measurements of the same physical quantity show a general pattern: the normal distribution grossly underestimates the probability of large deviations from the true values; the probability is well described by exponential functions. By analogy with physical measurements, the results indicate that the usual 95% confidence intervals in epidemiology and environmental studies should be expanded to account for unsuspected systematic errors.

1 Introduction

In a physical measurement such as length, published values are averages over many separate readings. If the estimates of the uncertainties are good, then by the Central Limit Theorem (CLT) the distribution around the true value will be asymptotically normal, or Gaussian. Uncertainty in a physical quantity can therefore be presented as an average of measurements, A, and an associated standard error, Δ. If the actual value is a then we expect $x = (a - A)/\Delta$ to be normal as well where x is the normalized deviation. The range $A \pm 1.96\Delta$ has a 95% probability of including a. The presence of systematic uncertainties, however, violates the assumptions necessary for use of the CLT. If most of the uncertainty comes from systematic errors, Gaussian distribution is no longer justified. Moreover, it is well known that researchers tend to underestimate the systematic uncertainties in their results [1]. Despite this, the normal distribution often remains implicit when researchers report measured values and the corresponding uncertainties [2].

The range of systematic uncertainty is not the standard deviation of a specified distribution, because the major uncertainties involved (e.g. detector efficiency) are not stochastic. In practice, the range of systematic uncertainty is often split in half and used as a surrogate standard deviation of the Gaussian distribution which is then added in quadrature with the estimated stochastic errors. The range reported by the authors represents their subjective judgement about the probability that the "true" value will lie within the specified range. Therefore uncertainty estimates are not "confidence intervals" in the classical statistical sense and are sometimes referred to as "subjective confidence intervals" [3]. The appropriate interpretation of the 95% confidence intervals in physical measurements is that there is a 95% chance that the interval will include the true value. A comparison of the empirical frequency of large deviations from the predicted values with normal distribution allows an analogy with much better understood stochastic uncertainties.

In this paper we analyze the time trends in five large nuclear data sets and describe a practical approach which utilizes these data in estimating the range of uncertainty in the individual physical measurements, observational and environmental studies and their meta-analytic syntheses.

2 Approach

The first attempts to quantify overconfidence in physical measurements come from the work of Bukhvostov [4] and Henrion and Fischoff [5] who analyzed the record of elementary particle data properties and fundamental constants while searching for occurrences of new measurements more than three

0-8186-3850-8/93 $3.00 © 1993 IEEE

standard deviations away from the previous values. They compared several hundred experimental results from early compilations with the "exact" values taken from a more recent compilation. A convenient measure of the deviation of "new" values from the "old" values is the normalized deviation $x = (a - A)/\Delta$, with a the exact value, A the measured value, and Δ the old standard deviation. The resulting distribution of $|x|$ had a tail extending beyond five standard deviations. By contrast, the cumulative normal distribution, $\mathrm{erfc}(x/\sqrt{2})$, estimates the probability of such fluctuation at only $5.6 \cdot 10^{-7}$.

In this paper we expand these original studies by following trends in five data sets: masses and lifetimes of elementary particles maintained by the Lawrence Berkeley Laboratory (LBL) Particle Data Group [6], magnetic moments [7] and lifetimes [8] of the excited nuclear states, neutron scattering lengths [9], and neutron resonance parameters [10,11].

3 Data Sets and Analysis Procedure

All data sets were first converted into a standard format. Each measurement that produced an experimental value, A, *and* an estimate of uncertainty, Δ, together with the date of publication (or incorporation into an electronic data bank) formed a separate line (record). Successive measurements of the same quantity comprised a block of data; a data set typically consisted of several hundred such blocks.

In order to limit the effects of "noise" in the data on final results, two selection criteria were applied.

i) The ratio of the old stated error, Δ_{old}, to the new error, Δ_{new}, had to be appreciable: $\Delta_{old}/\Delta_{new} \geq r$. For large value of r the difference between the new value, A_{new}, and the "true" value, a, was much less than Δ_{old}; however, this reduced the size of the final data set. The value r=4.0 was used; since errors are combined in quadrature: $\Delta_{tot}=(\Delta_{old}^2+\Delta_{new}^2)^{1/2}$. This changed the resulting x values by no more than 3%.

ii) We considered only those blocks for which the deviation from the true value was not too large ($|x| < m$; m=10 was used). This ensured that no major mistake occurred in the old measurement, e.g. that the nuclear excited state was correctly identified at that time. Thus we are excluding major errors.

Our selection procedure dramatically reduced the number of blocks remaining in each data set: from 124 to 79 for particle data; from 805 to 185 for nuclear moments; from 1691 to 214 for nuclear lifetimes; from 288 to 76 for neutron scattering; from 1203 to 62 for neutron resonance parameters. The remaining data were stable with respect to variations in the values of the parameters r and m. All of these sets were analyzed. For each block $x=(A_{old}-A_{new})/\Delta_{old}$ was calculated and empirical probabilities of $|x|$ were derived.

4 Parametrization of the Observed Distributions

The cumulative probability distributions of $|x|$ for each of the five data sets are also shown in Figure 1. They confirm the earlier findings [4,5,12,13] that a normal distribution grossly underestimates probability of large deviations from the expected values. A striking result of our analysis is that there is a general pattern in the empirical cumulative probability distributions of uncertainties of very different kinds of measurements: all distributions shown in Figure 1 are well fit by simple exponential functions with only one free parameter, u, which can be interpreted as a measure of overconfidence. A similar pattern with u values 3-4 is observed for energy and population projections [12-16].

To illustrate how the exponential functions can arise (see [14] for details) consider a set of estimates in which the mean and standard deviation are calculated by the standard method. Assume that the mean, A, is unbiased but that the estimate of the standard deviation, Δ, is randomly biased by systematic errors with a distribution $f(t)$. The distribution for $x = (a - A)/\Delta$ is then no longer a simple Gaussian, but can be written as a compound distribution $p(x) \propto \int_0^\infty f(t) exp[-x^2/2t^2]dt/t$. It appears that if $f(t)$ is sufficiently broad so that for large t, $f(t) \propto exp(-t^2/2u^2)$ then we find that $p(x) \propto exp(-|x|/u)$. The new parameter, u, is the relative uncertainty in the original standard deviation, Δ.

The normal ($u = 0$) and exponential ($u > 1$) distributions are members of a single-parameter family of curves shown in Figure 2. In this framework the parametric uncertainty can be quantified by analyzing the record of prior projections and estimating the value of u. The cumulative probability functions for $u \geq 1$, $x \geq 3$, can be approximated by $e^{-|x|/(0.7u+0.6)}$.

5 Estimating 95% Confidence Intervals in Observational Studies

Normal (Gaussian) approximations to the discrete distributions of the observed numbers (e.g. binomial) are common in epidemiology and results are usually presented in the form of 95 % confidence intervals. The general problem with observational studies is that numerous sources of bias are only taken into account on the basis of plausible assumptions which cannot be independently tested for the population under study. As Figure 1 shows, the usual 95 % confidence intervals, A ± 1.96Δ, estimated on the basis of early measurements miss the recent ("true") values not in 5% of cases as expected but in 20% of cases. This means that the common methods of calculating the confidence intervals should be modified to account for the *unknown* systematic uncertainties.

For example, the approximate confidence limits of the excess relative risk (usually called the incidence rate ratio), RR, are given by the formula [17].

$$RR = e^{\ln(RR) \pm Z \cdot SD(\ln(RR))}$$

where a value for Z is arbitrarily selected to give the desired confidence level for the normal distribution, e.g. Z=1.96 for 95 % confidence interval. Here a logarithmic transformation is used to compensate for the asymmetric sampling distribution of rate ratios.

If one assumes that the effects of incorrect classification and unaccounted for confounding factors in observational studies are no less important than the analogous effects of unknown systematic uncertainties in physical measurements, the new confidence limits can be determined from Figure 2. In other words, we assume that the relative "uncertainty of uncertainty" measured by the parameter u is the same in both cases. For u=0 (corresponding to normal distribution), the Z value for a 95 % confidence interval is 1.96 and for u=1 (roughly corresponding to the data set of physical measurements), Z=3.8. This means that many weak "statistically significant" findings in observational studies actually should no longer be regarded as significant.

It is evident that this procedure is valid when statistical sampling errors are not dominant. In the closed cohort studies, the epidemiologist identifies a set of subjects that existed in the past and follows over a specified time duration to determine their eventual fate. The statistical error in such studies often far exceeds the systematic error and the considerations of this paper may not apply. However cohort studies are very expensive and the majority of studies in epidemiology are case-control studies. In case-control studies the number of subjects with disease ("cases") and without disease ("controls") is compared with the number of these subjects exposed and not exposed to the factor in question. In case-control studies systematic errors are usually greater than in cohort studies. This paper is particularly concerned with the systematic errors that occur in these epidemiological methods.

The implications of the modified confidence intervals are illustrated in Figure 3. Here the results of 15 epidemiologic studies (13 case-control and 2 cohort) of the effects of the occupational exposure to electromagnetic fields on the risk of cancer (described in [18]) are summarized. Only those studies where the observed risk ratios, RR, exceeded 2 were included. With Z=1.96, the lower bound of the 95 % confidence intervals lie above RR=1 for 10 out of 15 studies making those studies "statistically significant". However with Z=3.8, the lower bounds of the expanded confidence intervals lie below RR=1 in all cases.

6 Estimating 95% Confidence Intervals for Lognormal Distributions

The lognormal distribution is widely used for describing probability distributions of variables that are essentially positive [19]. Its major advantage is that for the product of several factors uncertainties from different sources can be easily combined without the need for extensive computations. In particular, the lognormal distribution is used for uncertainty analysis of radiation risks [20,21].

If the logarithm of an estimate follows a normal distribution with mean, m, and standard deviation, s, then the estimated value follows a lognormal distribution characterized by its median $M=e^m$, and by its geometric standard deviation (GSD), $S=e^s$. Confidence intervals are calculated using the uncertainty factor, K, which is defined as $K=e^{\pm Z \cdot s}$. For 95% confidence intervals, Z=1.96 and there is 95% chance that the true value lies between $(1/K)\cdot M$ and $K \cdot M$.

The results of the analysis of the data sets of physical measurements indicating that normal distribution underestimates the probability of large deviations fully apply to lognormal distribution as well. Instead of Z=1.96 one should use Z=3.8 which results in much wider confidence intervals. For example, the NRC report [3] estimated the 95% confidence limits for the coefficient, C, relating the deep-dose equivalent (rem) to the film badge exposure worn by participants in each of the atmospheric nuclear tests series. To be specific, let us consider the doses for the Operation PLUMBBOB. Their best estimate of C (when the film badge exposure exceeded 0.2 Roentgen) was 0.77 with 95% confidence limits 0.51 to 1.15. Applying Z=3.8 gives new limits 0.35 to 1.68 resulting in more than doubling the range of uncertainty in the dose estimates. It seems worthwhile to perform this type of re-analysis for all other uncertainty estimates relating to the film badge dosimetry in course of the recently commissioned follow-up study of the same cohort [22]

7 Testing the Recipes of Meta-Analysis

Databases of physical measurements can be used to test the validity of the widely used procedure of calculating weighted averages over several measurements. More generally, this procedure is a subject of meta-analysis. The latter is a common name for numerical summaries of the results of multiple studies of the same effect which hopefully can offer better precision than the component studies [23]. We shall address these issues by treating successive measurements of one and the same parameter as independent component studies (see [24] for more details).

One problem with using weighted averages is that there may be true heterogeneity in the data. Although the underlying physical quantity is obviously the same for all studies, systematic errors could cause biased parameter estimates and hence contribute to the heterogeneity of the results. To this end Laird and Mosteller [25] developed a method specifically designed to take heterogeneity into account by including an additional variation component representing heterogeneity which adds up in quadrature with the intra-study variation. The results of testing the Laird-Mosteller approach against the elementary particle data set (LBL) are shown in Figure 4; similar results were obtained for other data sets. Taking the interstudy variation into account reduces the frequency of large $|x|$ values, however, the curves for averaging over 5 and 10 experiments still lie far above the curve for 1 experiment, not to mention the Gaussian curve. This indicates that the Laird-Mosteller algorithm indeed improves the description of the probability of large deviations but not nearly enough to offset the negative effect of combining the disparate studies.

8 Discussion of the Results

We have demonstrated that there is a general pattern in the empirical cumulative probability distributions of deviations from the old values in very different kinds of nuclear measurements. All distributions shown in Figure 1 can be fit by simple exponential functions with one free parameter, the ratio of the unsuspected systematic errors to the recognized uncertainties. For nuclear data $u \sim 1$. Our analysis provides a novel recipe for calculating 95% confidence intervals based on observed trends in past data.

This study shows that the standard uncertainty analysis must be supplemented with an additional step: analysis of "uncertainty of uncertainty", comparison of the unsuspected errors against the known uncertainties. Standard uncertainty analysis provides an estimate of the width of the probability distribution around the simple point estimate. However, the commonly used 95% confidence intervals are determined by the tails of the distribution that are extremely sensitive to underestimation of the uncertainties ("overconfidence"). Analysis of trends in past data allows to quantify the degree of overconfidence by fitting the empirical distribution and estimating the value of parameter u for each type of data.

The results of this study indicate that the commonly used criteria of statistical significance in observational and environmental studies must be revised. Measurements of fundamental physical constants are much better defined than epidemiological studies, yet overconfidence about their accuracy is widespread among experts [1]. There is every reason to expect that environmental and epidemiological studies incorporate no less unsuspected systematic error than the measurements in nuclear physics. Analysis of these measurements gives rise to tails in the probability distribution; we parametrize these tails by an exponential with $u=1$. The effects of systematic uncertainties are usually hidden in epidemiology because calendar time is an important parameter in itself. Moreover, due to the inherent variability of the system under study the "true" values may not exist.

Our analysis also shows that the common recipes for combining several experimental results fail to produce correct confidence intervals for physical measurements (where the exact answer is known). Empirical probability distributions of large deviations from the predicted values suggest that much wider confidence intervals should be used in order to capture 95 % probability both in the individual studies and their meta-analytic syntheses.

Acknowledgements

This research was partially supported by the U.S. Department of Energy through the Northeast Center of the National Institute for Global Environmental Change (NIGEC) and by the Biological Effects Branch of the U.S. Department of Air Force through the contract F33615-92-C-0602. We thank D.M. Kammen, N. Laird, S.Sakharov, A. Serebrov, O. Sumbaev, M. Stoto, and B. Yerozolimsky for discussion and comments; I. Kondurov and E. Seymann for sending their data sets on diskettes; E. Jin for computer assistance. Electronic mail addresses:
shlyakhter@physics.harvard.edu.
wilson@physics.harvard.edu

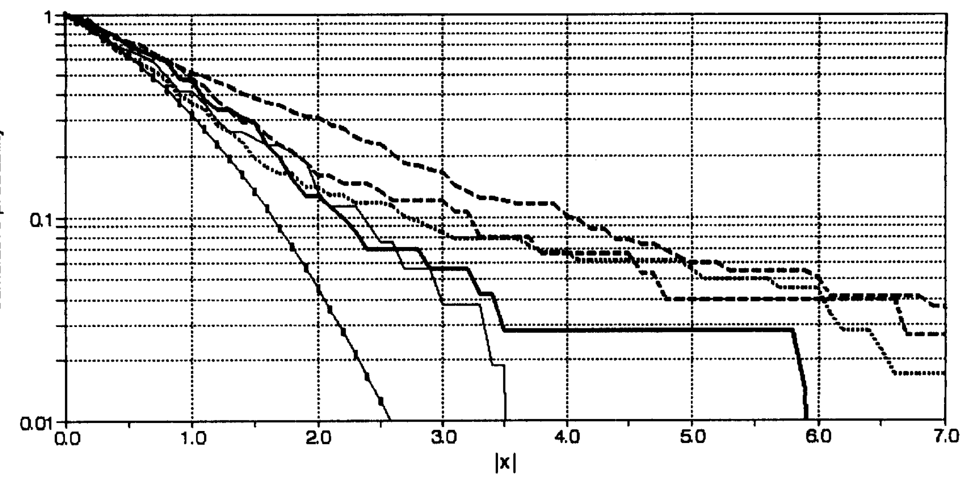

Figure 1: Probability of unexpected results in physical measurements. The plots depict the cumulative probability, $S(x)=\int_x^{\infty} p(t)dt$, that new measurements (a) will be at least |x| standard deviations (Δ) away from the old results (A); x = (a - A)/Δ as defined in the text. The cumulative probability distributions of |x| is shown for the five data sets: particle data [6] (heavy solid line); magnetic moments [7] and lifetimes [8] of excited nuclear states (respectively heavy centered line and heavy dotted line), neutron scattering lengths [9] (heavy dashed line), and average neutron resonance parameters [10,11] (solid line).

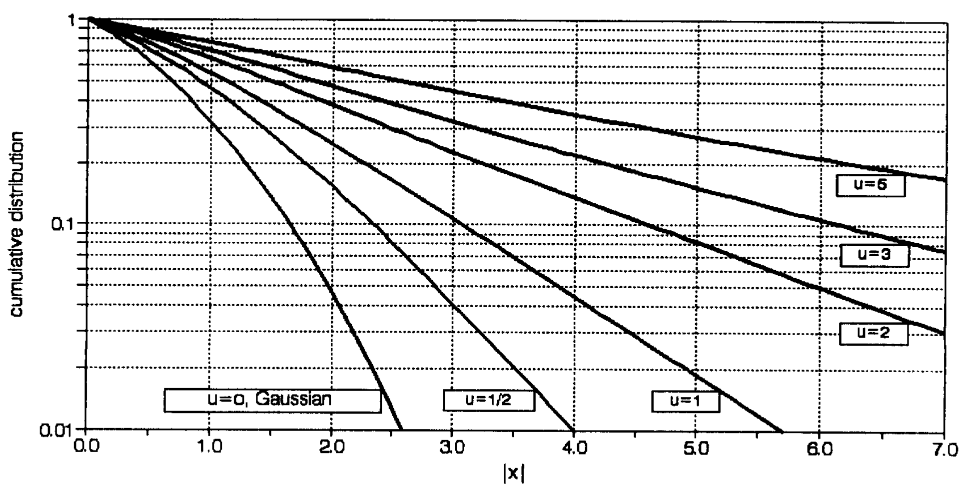

Figure 2: One-parameter set of probability distributions of deviations: parameter u defines the uncertainty in the standard deviation of the Gaussian distribution. The values of u are indicated in the figure. The curves demonstrate the continuum of probability distributions: from Gaussian (u=0) to exponential (u > 1).

and fundamental constants while searching for occurrences of new measurements more than three standard deviations away from the previous values. They compared several hundred experimental results from early compilations with the "exact" values taken from a more recent compilation. A convenient measure of the deviation of "new" values from the "old" values is the normalized deviation $x = (a - A)/\Delta$, with a the exact value, A the measured value, and Δ the old standard deviation. The resulting distribution of $|x|$ had a tail extending beyond five standard deviations. By contrast, the cumulative normal distribution, $\mathrm{erfc}(x/\sqrt{2})$, estimates the probability of such fluctuation at only $5.6 \cdot 10^{-7}$.

In this paper we expand these original studies by following trends in five data sets: masses and lifetimes of elementary particles maintained by the Lawrence Berkeley Laboratory (LBL) Particle Data Group [6], magnetic moments [7] and lifetimes [8] of the excited nuclear states, neutron scattering lengths [9], and neutron resonance parameters [10,11].

3 Data Sets and Analysis Procedure

All data sets were first converted into a standard format. Each measurement that produced an experimental value, A, *and* an estimate of uncertainty, Δ, together with the date of publication (or incorporation into an electronic data bank) formed a separate line (record). Successive measurements of the same quantity comprised a block of data; a data set typically consisted of several hundred such blocks.

In order to limit the effects of "noise" in the data on final results, two selection criteria were applied.

i) The ratio of the old stated error, Δ_{old}, to the new error, Δ_{new}, had to be appreciable: $\Delta_{old}/\Delta_{new} \geq r$. For large value of r the difference between the new value, A_{new}, and the "true" value, a, was much less than Δ_{old}; however, this reduced the size of the final data set. The value r=4.0 was used; since errors are combined in quadrature: $\Delta_{tot}=(\Delta_{old}^2+\Delta_{new}^2)^{1/2}$. This changed the resulting x values by no more than 3%.

ii) We considered only those blocks for which the deviation from the true value was not too large ($|x| < m$; m=10 was used). This ensured that no major mistake occurred in the old measurement, e.g. that the nuclear excited state was correctly identified at that time. Thus we are excluding major errors.

Our selection procedure dramatically reduced the number of blocks remaining in each data set: from 124 to 79 for particle data; from 805 to 185 for nuclear moments; from 1691 to 214 for nuclear lifetimes; from 288 to 76 for neutron scattering; from 1203 to 55 for neutron resonance parameters. The remaining data were stable with respect to variations in the values of the parameters r and m. All of these sets were analyzed. For each block $x=(A_{old}-A_{new})/\Delta_{old}$ was calculated and empirical probabilities of $|x|$ were derived.

4 Parametrization of the Observed Distributions

The cumulative probability distributions of $|x|$ for each of the five data sets are also shown in Figure 1. They confirm the earlier findings [4,5,12,13] that a normal distribution grossly underestimates probability of large deviations from the expected values. A striking result of our analysis is that there is a general pattern in the empirical cumulative probability distributions of uncertainties of very different kinds of measurements: all distributions shown in Figure 1 are well fit by simple exponential functions with only one free parameter, u, which can be interpreted as a measure of overconfidence. A similar pattern with u values 3-4 is observed for energy and population projections [12-16].

To illustrate how the exponential functions can arise (see [14] for details) consider a set of estimates in which the mean and standard deviation are calculated by the standard method. Assume that the mean, A, is unbiased but that the estimate of the standard deviation, Δ, is randomly biased by systematic errors with a distribution $f(t)$. The distribution for $x = (a - A)/\Delta$ is then no longer a simple Gaussian, but can be written as a compound distribution $p(x) \propto \int_0^\infty f(t) exp[-x^2/2t^2]dt/t$. It appears that if $f(t)$ is sufficiently broad so that for large t, $f(t) \propto exp(-t^2/2u^2)$ then we find that $p(x) \propto exp(-|x|/u)$. The new parameter, u, is the relative uncertainty in the original standard deviation, Δ.

The normal ($u = 0$) and exponential ($u > 1$) distributions are members of a single-parameter family of curves shown in Figure 2. In this framework the parametric uncertainty can be quantified by analyzing the record of prior projections and estimating the value of u. The cumulative probability functions for $u \geq 1$, $x \geq 3$, can be approximated by $e^{-|x|/(0.7u+0.6)}$.

5 Estimating 95% Confidence Intervals in Observational Studies

Normal (Gaussian) approximations to the discrete distributions of the observed numbers (e.g. binomial) are common in epidemiology and results are usually presented in the form of 95 % confidence intervals. The

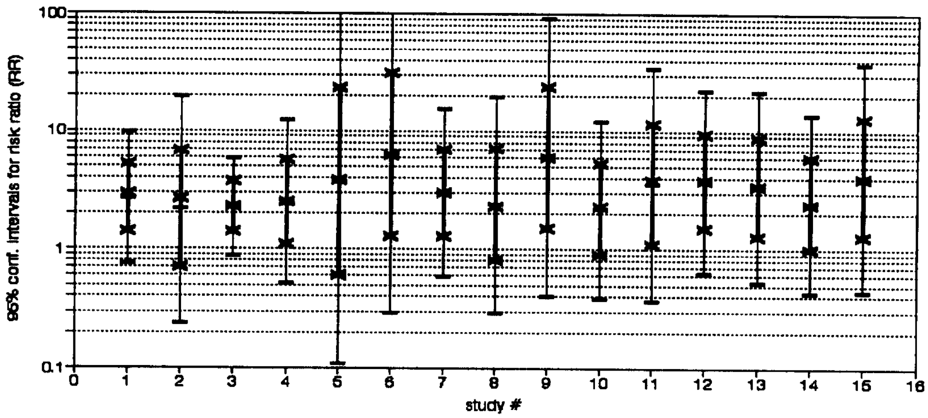

Figure 3: The implications of the wider confidence intervals for the results of 15 epidemiologic studies of the effects of the occupational exposure to electromagnetic fields on the risk of cancer. To illustrate the accuracy of the normal approximation to the exact Poisson confidence intervals, the center of the Poisson interval is marked by asterisk and the center of an equivalent normal interval is marked with a horizontal line. These markers almost coincide with each other for 14 out of 15 intervals.

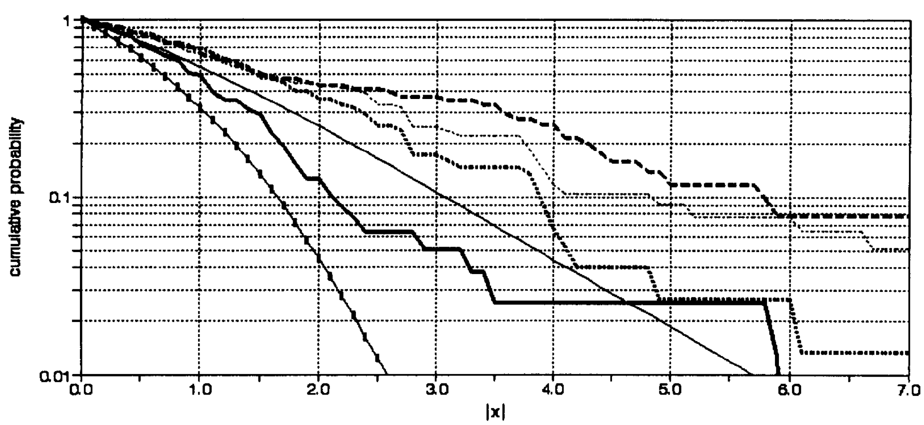

Figure 4: Testing the algorithms for calculating the weighted averages against the particle (LBL) data set: no averaging (1 experiment, heavy solid line, same as in Figure 1); standard weighted average over 5 experiments (dashed line); accounting for heterogeneity (systematic errors and inter-study variation) [20] for 5 (heavy dashed line) and 10 (heavy dotted line) experiments. For comparison the curve u=1 from Figure 2 is also shown.

References

[1] M.G. Morgan and M. Henrion *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis* (Cambridge University Press, New York, 1990).

[2] C.F. Dietrich, *Uncertainty, Calibration and Probability: The Statistics of Scientific and Industrial Measurement*, (Adam Hilger: Bristol UK, 1991).

[3] National Research Council *Film badge dosimetry in atmospheric nuclear tests*, National Academy Press, Washington, D.C. 1989.

[4] A.P. Bukhvostov, "On the statistical treatment of experimental data," Preprint, Leningrad Nuclear Physics Institute LNPI-45 (1973) (in Russian).

[5] M. Henrion and B. Fischoff *American Journal of Physics* 54, 791 - 797 (1986).

[6] A.H. Rosenfeld, *et al., Rev. Mod. Physics*, 36, 977-1004, 1964.

[7] M.P. Avotina, I.A. Kondurov, and O.N. Sbitneva "Tables of nuclear moments and deformation parameters of atomic nuclei", Leningrad Nuclear Physics Institute Report, 1982.

[8] E.E. Berlovich, L.A. Vaishnene, I.A. Kondurov, Yu.N. Novikov, and Yu.V. Sergeenkov (1975) "Tables of lifetimes of nuclear levels", Leningrad Nuclear Physics Institute Report.

[9] L. Koester, H. Rauch, and E. Seymann (1991) "Neutron scattering lengths: a survey of experimental data and methods", *Atomic Data and Nuclear Data Tables*, 49:65-120.

[10] S.F. Mughabhab and D.I. Garber (1973) *Neutron Cross Sections* Vol.1, "Resonance Parameters," 3rd ed., BNL-325, Brookhaven National Laboratory.

[11] S.F. Mughabhab *et al.* (1981) "Neutron Cross Sections," Vol. 1, *Neutron Resonance Parameters and Thermal Cross Sections*, Part A, Academic Press, New York.

[12] A.I. Shlyakhter and D.M. Kammen, "Sea-level rise or fall?" *Nature*, 357, 25, 1992.

[13] A.I. Shlyakhter and D.M. Kammen "Estimating the range of uncertainty in future development from trends in physical constants and predictions of global change," CSIA discussion paper 92-06, Kennedy School of Government, Harvard University, July 1992

[14] A.I. Shlyakhter and D.M. Kammen "Uncertainties in Modeling Low Probability/High Consequence Events: Application to Population Projections and Models of Sea-level Rise", these proceedings.

[15] D.M. Kammen, A.I. Shlyakhter, C.L. Broido and R. Wilson "Non-Gaussian Uncertainty Distributions: Historical trends and Forecasts of the United States Energy Sector, 1983-2010," these proceedings.

[16] A.I. Shlyakhter, D.M. Kammen, C.L. Broido, and R. Wilson, "Quantifying the Credibility of Energy Projections from Trends in Past Data: the U.S. Energy Sector," paper submitted to *Energy Policy*.

[17] K.J. Rothman (1986) *Modern Epidemiology*, Little, Brown & Co., Boston.

[18] *Health Effects of Low-Frequency Electric and Magnetic Fields*, Report of the Oak Ridge Associated Universities Panel for the Committee on Interagency Radiation Research and Policy Coordination ORAU 92-F8, Washington, D.C., 1992, 372p.

[19] E.L. Crow and K. Shimizu (1988), *Lognormal Distribution. Theory and Applications*, Marcel Dekker, Inc., New York and Basel.

[20] National Council on Radiation Protection and Measurement (1984) "Radiological assessment: predicting the transport, bioaccumulation and uptake by man of radionuclides released to the environment", Washington, DC:NCRP; Report No.76.

[21] National Institutes of Health (1985) "Report of the NIH *ad hoc* working group to develop radioepidemiological tables", Washington, DC: US Government Printing Office; NIH Publication No. 85-2738.

[22] C. Anderson "NAS to redo atomic studies found to be flawed", *Nature* (News), 359, 354, 1992.

[23] L.V. Hedges and I. Olkin *Statistical methods for meta-analysis"*, Academic Press Inc., 1985.

[24] A.I. Shlyakhter and R. Wilson, "Estimating uncertainty and combining data sets in observational studies," paper in preparation.

[25] N.M. Laird and F. Mosteller, "Some statistical methods for combining experimental results," *International Journal of Technology Assessment in Health Care*, 6:5-30, 1990.