Bayesian Inference and Maximum Entropy
Methods in Science and Engineering
K. Knuth et al. eds, AIP Conf. Proc. (2005)

LA-UR-05-6743

Bayesian Analysis of Inconsistent Measurements of Neutron Cross Sections

Kenneth M. Hanson

Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545
kmh@lanl.gov

Abstract. The evaluation of neutron cross sections as a function of energy is fraught with inconsistent measurements. I describe a Bayesian approach to deal with the inconsistencies by probabilistically modeling the possibility of discrepant data and data sets with long-tailed likelihood functions. Systematic normalization uncertainties in each data set are included by considering the normalization to be a variable with specified uncertainty. By characterizing its uncertainty with a mixture of Cauchy and Gaussian distributions, data sets that disagree with the majority of others are given less weight in terms of normalization, but still provide useful information about the energy dependency of the cross sections. I demonstrate the approach with data sets of neutron fission cross sections for americium 243. Samples from the posterior obtained with the Markov Chain Monte Carlo technique are used to estimate the posterior mean and standard error.

Keywords: neutron fission cross sections, americium-243, Bayesian analysis, normalization uncertainty, discrepant data, outliers, long-tailed likelihood, Cauchy distribution, Student t-distribution
PACS: 02.50.Tt, 02.50.Ng, 02.50.Sk, 06.20.Dk, 07.05.Kf, 07.05.Tp, 25.85.Ec

INTRODUCTION

The process of evaluation of neutron cross sections involves combining measurements from numerous experiments to obtain a single set of cross-section value as a function of energy and estimate their uncertainties. The experiments span many decades and are of varying quality. It is not unusual for the evaluator to be confronted with reported cross sections that are in substantial disagreement. It is important to take into account the independent (random) errors in each datum, as well as the systematic uncertainty in the normalization of each experiment. Regrettably, the published data are not always accompanied with detailed description of the sources of their uncertainties.

The present work focuses on a Bayesian approach to cope with inconsistent measurements and discrepant data sets. A necessary ingredient in the analysis is the assignment of uncertainty in the data, both in the form of independent errors and normalization error. The latter is treated as a systematic effect common to all data from a single experiment. I develop a probabilistic model of the measurements, which includes these separate sources of uncertainty. The underlying error distributions are taken to be Gaussian. However, if the nominal width assumed for the error distribution is considered to be uncertain, and underestimated, the resulting likelihood function should possess a long tail, which is known to ameliorate the effect of outliers. The final estimates for the neutron cross sections are obtained by sampling the posterior distribution by means of Markov Chain Monte Carlo (MCMC) to obtain their means and variances. I demonstrate the approach on measurements of the neutron fission cross sections for americium-243.

BAYESIAN TREATMENT OF DISCREPANT DATA

A measurement that disagrees with numerous other data by much more than its stated standard error is called an outlier. Outliers are often caused by mistakes made in taking the data or their analysis. While we might prefer to believe otherwise, every experimentalist knows that mistakes happen. Of course, experience and carefulness can reduce the number of mistakes, but not eliminate them entirely. For dramatic evidence of outliers, look at the review of the properties of fundamental particles periodically performed by the Particle Data Group (2004) [1]. The question is, how do we cope with outliers?

The traditional approach to dealing with outliers is to identify and eliminate them from the analysis. In this iterative process, sometimes referred to as robust estimation, it may be difficult to decide which data are outliers. The data are typically either in or out. The Bayesian approach is to explicitly allow for the possibility of outliers by using long-tailed likelihood functions. The analysis is essentially automatic and includes all the data. The weight of each datum in determining the posterior mean is regulated by how well it agrees with the remaining data.

O'Hagan [2] outlined the early history of handling outliers in Bayesian analysis, which stretches back to 1961. He refined the proof of Dawid [3] that a likelihood function with a "thick" tail effectively reduces the influence of highly discrepant data points. An approach to treating discrepant measurements was presented at the MaxEnt meeting in 1988 by Fröhner [4] in which he estimated the "unrecognized errors" in each datum. In 1993 Hanson and Wolf [5] noted the potential usefulness of the Cauchy distribution for handling outliers and presented analytic expressions for the posterior mean and variance for Cauchy likelihood functions. Sivia [6] proposed a "good and bad data" approach in 1996, which leads to a mixture of two Gaussians. Press [7] argued similarly but included the probability that each datum belonged to the "good" distribution. Dose and von der Linden [8] proposed a gamma distribution for the prior on the inverse variance, under the assumption that the analyst could estimate of how accurately the experimenter has assessed his/her uncertainties. The result is a Student t-distribution for the likelihood function. In the context of a good-bad datum strategy, they estimated the probability that each datum came from a Gaussian with its quoted σ.

Uncertain uncertainties

The Gaussian distribution is often the appropriate form for a likelihood function when the standard error σ is known. However, when the data disagree by much more than their standard errors, we must conclude that their quoted uncertainties are probably incorrect. In that case, it is appropriate to scale each quoted σ_i by a factor s_i. The initial uncertainty in the vector $\mathbf{s}$ is characterized by the prior distribution $p(\mathbf{s}\,|\,I)$, where I stands for whatever information can be brought to bear. The posterior for the inferred quantity x is obtained by marginalizing the joint distribution for x and $\mathbf{s}$ over $\mathbf{s}$ [9]

$$p(x\,|\,\mathbf{d},\boldsymbol{\sigma}) = \int p(x,\mathbf{s}\,|\,\mathbf{d},\boldsymbol{\sigma})\mathrm{d}\mathbf{s} \propto \int p(\mathbf{d}\,|\,x,\mathbf{s},\boldsymbol{\sigma})p(x,\mathbf{s}\,|\,\boldsymbol{\sigma})\mathrm{d}\mathbf{s}\,, \tag{1}$$

where $\mathbf{d}$ is the vector of measurements and $\boldsymbol{\sigma}$ represents their stated standard errors. Under the assumption of an underlying Gaussian likelihood, a flat prior on x, and that the prior on s_i is independent of σ_i, the posterior is

$$p(x\,|\,\mathbf{d},\boldsymbol{\sigma}) \propto \int p(\mathbf{d}\,|\,x,\mathbf{s},\boldsymbol{\sigma})p(\mathbf{s})\mathrm{d}\mathbf{s} \propto \int p(\mathbf{s})\prod_i s_i^{-1}\sigma_i^{-1}\exp\left\{-\sum_i \frac{(x-d_i)^2}{2s_i^2\sigma_i^2}\right\}\mathrm{d}\mathbf{s}\,. \quad (2)$$

For certain choices of $p(s_i)$, the integration over s_i can be done analytically. The result is a revised likelihood function that includes the uncertainty in the standard error for each datum.

The various approaches for coping with outliers mentioned above differ mostly in the assumptions made for $p(\mathbf{s})$. The simplest assumption is that the measurements are drawn from two distributions, one corresponding to $s = 1$, and another with a much larger standard deviation, $s = \gamma \gg 1$. The first term corresponds to the well-behaved (good) measurements, the second to the outliers (bad). The resulting likelihood function for a measurement d, given a value x, is a mixture of two Gaussians [6, 7] (2G)

$$p(d\,|\,x) = \frac{(1-\beta)}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-d)^2}{2\sigma^2}\right\} + \frac{\beta}{\gamma\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-d)^2}{2\gamma^2\sigma^2}\right\}\,, \quad (3)$$

where β is the probability of d being an outlier.

As mentioned earlier, Dose and von der Linden [8] proposed a gamma distribution for $p(s^{-2}) = \Gamma(s^2, a)$, which yields a Student t-distribution for the likelihood. This prior is the same as the often-used "inverse-chi-squared" distribution [10]. The variable a is related to our prior belief about the uncertainty in s, $a^{-1} = \mathrm{var}(s^{-2})$. For the choice $a = \frac{1}{2}$, the likelihood is a Cauchy distribution. For that case, if one believes that a portion of the data comes from a Gaussian likelihood with the quoted σ, but admits the possibility of an outlier, the likelihood is a mixture of Cauchy and Gaussian distributions (CG)

$$p(d\,|\,x) = \frac{(1-\beta)}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-d)^2)}{2\sigma^2}\right\} + \frac{\beta}{\sigma\pi\sqrt{2}}\left(1+\frac{(x-d)^2}{2\sigma^2}\right)^{-1}\,. \quad (4)$$

Dose and von der Linden further considered β to be an uncertain variable, and therefore put a distribution on it and integrated over it to obtain the posterior on x.

It can be useful to draw an analogy between minus the logarithm of a probability density function ($= \varphi$) and a physical potential. Figure 1 shows the φ's for a Gaussian, a two-Gaussian mixture (3), and a Cauchy-Gaussian mixture (4). If φ is a potential, then its gradient $\nabla\varphi$ is the force with which a datum d pulls on the inferential quantity x. The effective force for a Gaussian increases linearly with the difference $|x-d|$. On the other hand, for the other two functions, which have long tails, the restoring force eventually decreases for large residuals. The behavior of $\nabla\varphi$ explains why these functions are tolerant of outliers. The gradient $\nabla\varphi$ corresponds to the influence function, employed in robust estimation to characterize the influence a new datum will have on an estimated quantity.

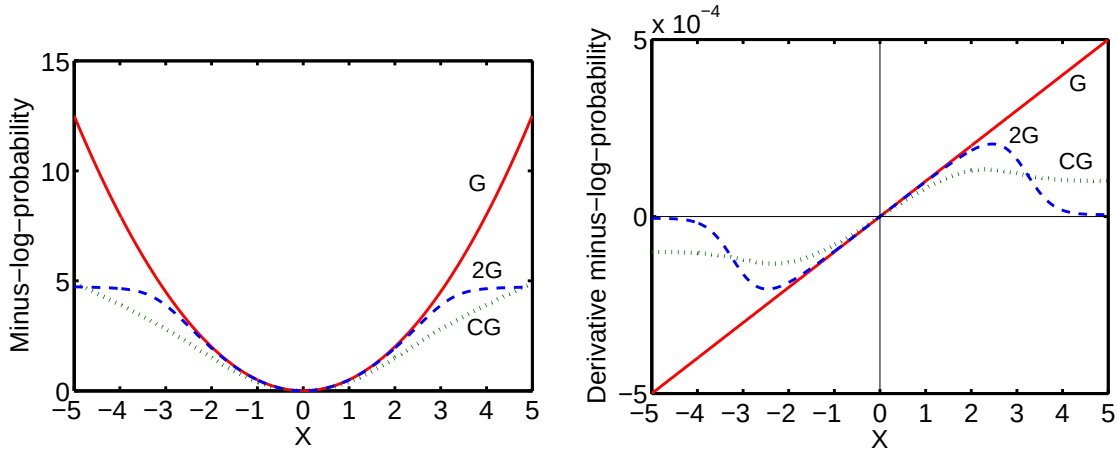

FIGURE 1. (a, left) Plot of φ (minus-log-probability) for several pdfs: G stands for Gaussian, 2G for a mixture of two Gaussians ($\beta = 0.1, \gamma = 10$), and CG for Cauchy-Gaussian mixture ($\beta = 2/3$). (b, right) the derivative of (a). Because φ is analogous to a potential, its derivative $\nabla\varphi$ acts like a force. The common feature of outlier-tolerant pdfs is that $\nabla\varphi$ falls off for large residuals.

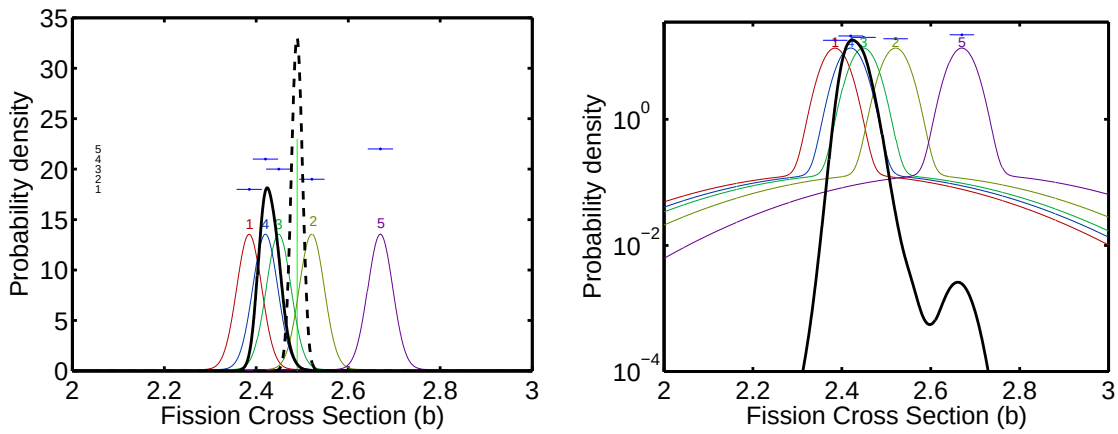

FIGURE 2. (a, left) When combined using Gaussian likelihoods, the five hypothetical measurements shown yield the posterior plotted as the dashed line. This result seems implausibly narrow, given the large dispersion in the data. The two-Gaussian mixture ($\beta = 0.1, \gamma = 10$) yields a more sensible posterior (solid black line), which effectively discounts the two outliers. (b, right) Semi-log plot shows the tails of the likelihoods and the posterior.

Example of discrepant data

As an example, suppose five measurements of a cross section are provided; 2.385, 2.521, 2.449, 2.420, and 2.670 b. All are stated to have a standard error of 0.027 b. Figure 2 shows the data with their error bars and corresponding likelihoods. Data point 5 is clearly an outlier, being eight standard deviations away from the mean of the other four points. A least-squares fit to these data gives the estimate of 2.489 ± 0.012 b, with $\chi^2 = 69.9$ and a p value of 2×10^{-14} for 4 DOF. This standard error seems unreasonably

small because the data are inconsistent with the underlying assumption of a Gaussian likelihood.

A well-known result of the Gaussian likelihood is that the standard deviation σ of the weighted average is given by $\sigma^{-2} = \sum_i \sigma_i^{-2}$, where σ_i is the standard error in the ith measurement. This automatic decrease in uncertainties does not depend on the distribution of the measurements, which defies common sense. When the measurements are more widely distributed than their σ's imply, more thought should be given to whether the model being used for the likelihood is appropriate.

With a flat prior on the cross section x, the posterior is proportional to the product of likelihoods: $p(x\,|\,\mathbf{d}) \propto \prod_i p(d_i\,|\,x)$. With the two-Gaussian likelihood ($\beta = 0.1, \gamma = 10$), the posterior mean and rms deviation are 2.430 $\pm$ 0.022 b. This result is close to the cluster of three points, which is what we might expect if we believe that the outlier should be discounted. The outlier has little effect. The standard error is much larger than for the Gaussian likelihood, but this is plausible because the three points in the cluster are more disperse than their σ's might indicate.

The semi-log plot, Fig. 2b, shows the tails of the likelihood for each datum. The posterior distribution, which is decidedly non-Gaussian, has a second small peak near the outlying point. This bump has little effect on the posterior mean and variance. The presence in the posterior of more than one local maximum, and the fact that the posterior may not be symmetric, indicates the need to use MCMC [11] to sample the posterior and estimate its mean and standard deviation.

This example points out how a simple long-tailed likelihood function does a remarkable job of handling outliers. The common experience reported in the articles already mentioned is that the details of the likelihood function employed are relatively unimportant. The important ingredient is that the likelihood possesses a long enough tail to encompass any potential outlier. For example, with the two-Gaussian function used here, if the broad Gaussian does not reach the outlier, it fails to ameliorate its disruptive effect.

NORMALIZATION UNCERTAINTY

In the general experimental scheme, cross sections are obtained by multiplying an observed number of counts by a normalizing factor. In this section, I focus on how to handle uncertainties in the normalization factor. A reasonable model for the measured number of counts is $d = cx + \varepsilon$, where x stands for the cross sections, c is the normalization factor, and ε is a random variable representing the statistical fluctuation in d. The counting process follows a Poisson distribution, but most cross-section experiments involve enough counts that a Gaussian distribution for ε is a good approximation. Assuming normal distributions for the uncertainty in c and writing $p(cx, c\,|\,\mathbf{d}) \propto \exp(-\varphi)$, the negative logarithm of the posterior distribution for each data set is

$$2\varphi = \sum_i \frac{(cx - d_i)^2}{\sigma_i^2} + \frac{(c-1)^2}{\sigma_c^2}\ , \tag{5}$$

where σ_c is the relative uncertainty in the normalization factor.

We desire the distribution in the variables x and c, so we need to transform variables. Consequently, we must divide the $p(cx,c)$ by $|J|$, where J is the Jacobian for the transformation [9], i.e. the determinant of the matrix of first derivatives of the new variables with respect to the old ones; $J = 1/c$. Thus, $p(x,c) \propto |J|^{-1} \exp(-\varphi)$. To obtain $p(x)$, the nuisance parameter c is marginalized out by integrating $p(x,c)$ over c. The Jacobian factor makes little difference in the present examples because the posterior distributions are quite narrow. Likewise, the more appropriate choice of a log-normal distribution [12] for the prior on c is not important here. For multiple data sets, one multiplies the likelihoods together, or sums the φ_k's from all data sets: $\varphi = \sum_k \varphi_k$.

Direct probabilistic modeling of multiplicative normalization uncertainties has been shown to resolve the PPP problem by Smith [13] (p. 205ff) and Hanson et al. [12].

DISCREPANT DATA SETS

When neutron cross sections for a nucleus have been measured many times, there is often some disagreement among the various measurements. Unfortunately, there is often too little information provided about the uncertainties in the data, especially systematic uncertainties. Then, the analyst has to make do with the data and their stated uncertainties.

Four sets of measurements of neutron-induced fission cross sections for americium-243 in the energy range 1–4 MeV are shown in Fig. 3. The labels and normalization errors for these are Fursov-1985, 1.4% [14], Behrens-1981, 0.9% [15], Knitter-1985, 2.8% [16], and Kanda-1987, 1.8% [17]. The one standard deviation error bar for normalization of each data sets in shown on the graphs as think vertical bars at around 1.64 MeV.

The energy dependence of the cross sections is modeled using cubic splines with nine knots, evenly spaced in $\log(E)$. The normalization is treated as a systematic effect, as described in the previous section. The posterior is sampled using MCMC based on the Metropolis algorithm with 2×10^5 steps. Figure 3 shows the results of the analysis done using the conventional Gaussian likelihood for both the data and the normalization term, as in Eq. (5). When the three-data-set cluster is analyzed, one obtains the posterior mean shown in Fig. 3a and the standard deviation visualized in Fig. 3b. The latter plot demonstrates the Bayesian notion of "model checking" in which the posterior predictive distribution is compared to the original data to see how well it matches them. In this case, the agreement is fairly good. These results seem reasonable.

However, when the fourth discrepant data set (Behrens-1981) is included, the results go awry. The posterior mean lies between the outlying data set and the others, and the standard error is smaller than for the three data sets. Figure 3d shows that the posterior predictive distribution does not represent the data, a sign that the analysis model is faulty. The inferred values of the cross section at 2 MeV are 1.453 ± 0.021 b for three data sets, and 1.588 ± 0.016 b for four. The latter relative uncertainty is just 1.0% so the normalization uncertainty in the Behrens-1981 data (0.9%) plays a big role in determining the final uncertainty. This result is very unsettling, since the normalization of this data set is highly suspect.

Figure 4 shows the results for the same data sets obtained using the Cauchy-Gaussian

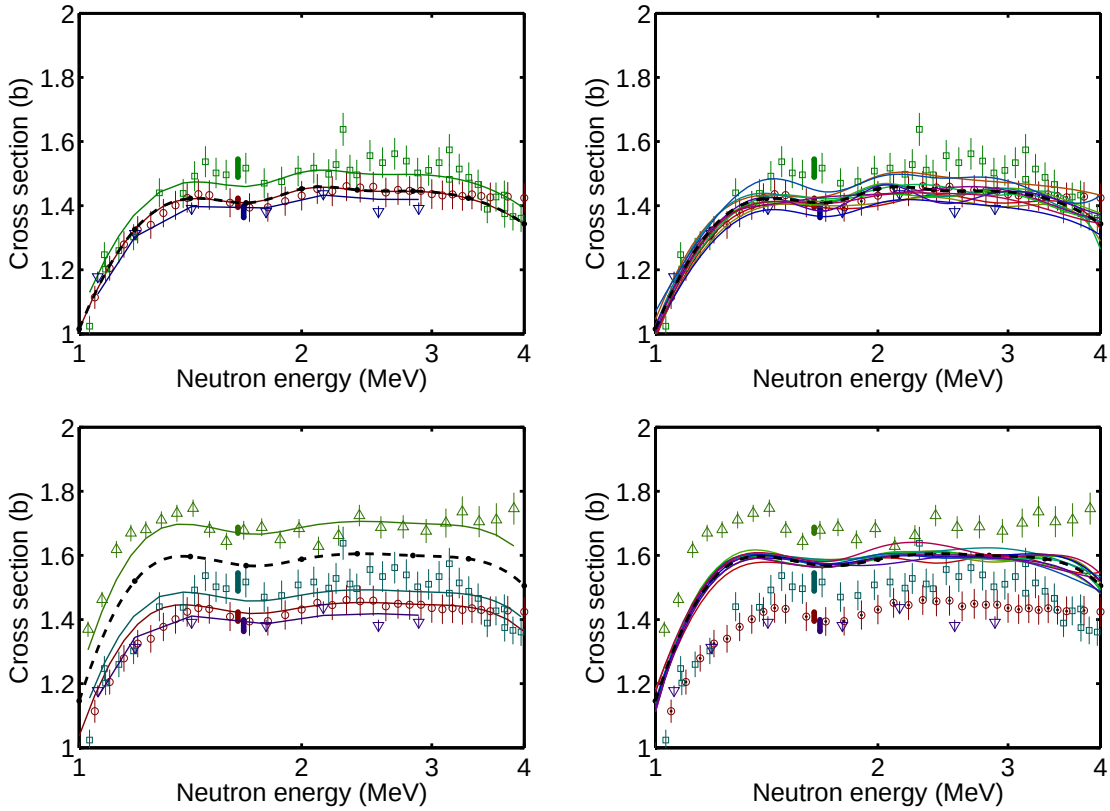

FIGURE 3. Analysis of fission cross sections for Am-243 based on Gaussian likelihood functions. The data sets are: ○ Fursov-1985, △ Behrens-1981, □ Knitter-1985, ▽ Kanda-1987. (a, top-left) Three data sets clustered together, with the dashed line showing the posterior mean. The other curves show the posterior renormalized to predict each data set. (b, top-right) 12 samples from posterior distribution compared to the data. (c, bottom-left) Same as (a) with a fourth discrepant data set added, which severely shifts the posterior mean. (d, bottom-right) Posterior samples for Gaussian analysis of all four data sets, which do not replicate any of the measurements.

mixture (4) with $\beta = 2/3$ for the normalization term in Eq. (5). This outlier-tolerant distribution allows for the possibility of a gross error in normalization. The results for the three data sets are not much different than obtained above, however, with slightly larger uncertainties. The cross sections at 2 MeV are 1.453 ± 0.026 b for three data sets, and 1.418 ± 0.021 b for four. Figures 4b and 4d show that the posterior predictive distributions match the three clustered data sets fairly well; the normalization of the fourth data set is effectively disregarded. The Cauchy-Gaussian mixture has the effect of increasing the standard error for the three-data-set cluster, which is not unreasonable, owing to their dispersion. It effectively reduces the influence of the Behrens-1981 data set. In fact, the cross section at 2 MeV is decreased a fair amount, but is increased elsewhere. The normalization of the outlying data set has no influence on result, but its shape is included, as evidenced by the bump in the posterior mean below 1.4 MeV.

I find that using a pure Cauchy distribution for the normalization term ($\beta = 1$), yields 1.460 ± 0.047 b and 1.449 ± 0.044 b for the two analyses. The width of the posterior is significantly increased by the long tails of the Cauchy. This choice amounts to using

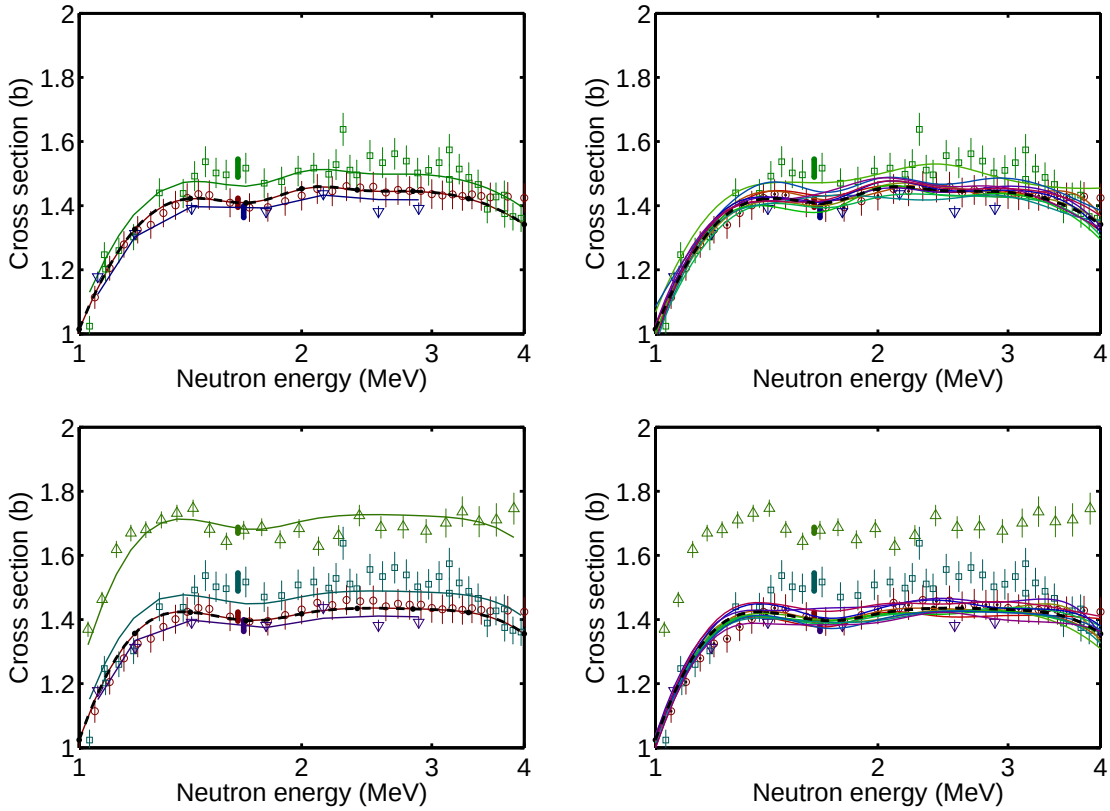

FIGURE 4. Analysis of the same data in Fig. 3. (a, top-left) The dashed line shows the posterior mean. The other curves show the renormalized posterior, which should match each data set. (b, top-right) 12 samples from posterior distribution compared to the data. (c, bottom-left) The posterior mean (dashed line) is hardly affected when the fourth discrepant data set is added. (d, bottom-right) Posterior samples for the Cauchy-Gaussian analysis of all four data sets.

a very broad prior for $p(\sigma)$. When β is smaller, its value is relatively unimportant for determining the posterior mean, but it has some effect on the standard deviation.

This analysis has not considered the potential outliers within each data set, but that is easy to do and might be warranted.

CONCLUSION

I have reviewed the general Bayesian approach to dealing with discrepant data, which is to use a long-tailed likelihood. When applied to the systematic normalization factor for a complete data set, the normalization factor is essentially ignored when the data set is discrepant. On the other hand, the energy dependence among the data belonging to that data set is incorporated into the posterior.

Bayesian analysis provides the logical and computational structure to allow one to combine knowledge about experiments in a consistent way. In a real sense, the Bayesian treatment of outliers presented here amounts to taking a vote among the data sets. The majority (weighted inversely by their variances) wins. It is, after all, possible that the

outlying data set could be the correct one. Therefore, the analyst must try to uncover all aspects of the experimental uncertainties and incorporate them into the analysis. Other kinds of independent measurements, such as criticality experiments, can be useful for resolving inconsistent data [18].

More work needs to be done on systematic uncertainties. It seems it would be beneficial to use more informative priors for the systematic uncertainties based on knowledge of the experiments, i.e., how they were done, the experimental techniques used, and who did them. A thorough analysis should be done on what kinds of uncertainties are typical and how to include them in a full probabilistic model of the experimental uncertainties.

ACKNOWLEDGMENTS

Toshihiko Kawano, Patrick Talou, and Gerry Hale have led me to a better understanding of the critical issues in cross-section evaluation. I am indebted to Toshihiko and Patrick for their critical review of this paper. This work was supported by the United States Department of Energy under contract W-7405-ENG-36.

REFERENCES

1. The review of physics particles (2004), Particle Data Group, http://pdg.lbl.gov.
2. A. O'Hagan, *J. Roy. Statist. Soc.* **B 41**, 358–367 (1979).
3. A. P. Dawid, *Biometrika* **60**, 664–667 (1973).
4. F. H. Fröhner, "Bayesian evaluation of discrepant experimental data," in *Maximum Entropy and Bayesian Methods*, edited by J. Skilling, Kluwer Academic, Dordrecht, 1989, pp. 467–474.
5. K. M. Hanson, and D. R. Wolf, "Estimators for the Cauchy distribution," in *Maximum Entropy and Bayesian Methods*, edited by G. R. Heidbreder, Kluwer Academic, 1996, pp. 255–263.
6. D. S. Sivia, "Dealing with duff data," in *Maximum Entropy and Bayesian Methods*, edited by M. Sears et al., N.M.B. Printers, Port Elizabeth, 1996, pp. 131–137.
7. W. H. Press, "Understanding data better with Bayesian and global statistical methods," in *Unsolved Problems in Astrophysics*, edited by J. N. Bahcall, and J. P. Ostriker, Princeton University, Princeton, 1997, pp. 49–60.
8. V. Dose, and W. von der Linden, "Outlier tolerant parameter estimation," in *Maximum Entropy and Bayesian Methods*, edited by W. von der Linden et al., Kluwer Academic, Dordrecht, 1999, pp. 47–56.
9. D. S. Sivia, *Data Analysis - A Bayesian Tutorial*, Clarendon, Oxford, 1996.
10. A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, Chapman & Hall, London, 1995.
11. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London, 1996.
12. K. M. Hanson, T. Kawano, and P. Talou, "Probabilistic interpretation of Peelle's pertinent puzzle and its resolution," in *Int. Conf. Nuclear Data for Science and Technology*, edited by R. C. Haight et al., AIP Conf. Proc. 769, 2005, pp. 304–307.
13. D. L. Smith, *Probability, Statistics, and Data Uncertainties in Nuclear Science and Technology*, vol. 8 of *OECD Nuclear Energy Agency Nuclear Data Committee Series, Neutron Physics and Nuclear Data in Science and Technology*, American Nuclear Society, LaGrange Park, 1991.
14. B. I. Fursov et al., *Atomnaya Energiya* **59**, 339–343 (1985).
15. J. W. Behrens, and J. C. Browne, *Nucl. Sci. Eng.* **77**, 444 (1981).
16. H. Knitter, and C. Budtz-Jorgensen, *Nucl. Sci. Eng.* **99**, 1 (1988).
17. K. Kanda et al., *J. Nucl. Sci. Tech.* **24**, 423–430 (1987).
18. T. Kawano et al., *to be published in Nucl. Sci. Eng.* (2005).