OUTLIER TOLERANT PARAMETER ESTIMATION

V. DOSE AND W. VON DER LINDEN
Max-Planck-Institut für Plasmaphysik, EURATOM Association
D-85740 Garching b. München, Germany ¶

Abstract. Real world does not only provide noisy instead of perfect data. Every experimentalist has now and then to deal with outliers. The situation is simple if isolated points stick out of the general trend by a large amount. Arguments can then usually be found why such a point should be disregarded. The situation becomes critical if the outliers are not that obvious. This is usually the case for parameter space dimensions ≥ 3. We present a Bayesian solution to the outlierproblem which assumes that the uncertainties assigned to the experimental data are only estimates of the true error variances.

Key words: Outlier, parameter estimation, duff data

1. Introduction

It is common experience to all experimental scientists that repeated measurements of supposedly one and the same quantity result occasionally in data which are in striking disagreement with all others. There are several conceivable reasons for such outliers. The apparatus may have performed differently due to some unstable component. If this instability has escaped the attention of the experimenter then there is no good reason to delete the outlying data point. The second possibility is that everything went well and the seemingly discordant data point has occurred as a highly unlikely event compatible with the sampling statistics. Finally, the discordant data point may signal a new and unexpected effect. This case is the most difficult and uninteresting in the context of the present paper since only repeated measurements will gradually fix the new phenomenon. By exactly this effort also the first two cases may be healed. So where is the problem? The problem is that only a small number of measurements is sufficiently fast, cheap and easy to perform such that the outlierproblem can be overcome just by increased measurement efforts. Everybody who has chosen this way at some time in his career knows what increased efforts means depending on the extent to which the outlying data point deviates from the mainstream.

¶e-mail: wvl@ipp-garching.mpg.de

The situation which we have considered so far is simple: a single outlierand at least in principle augmentable data. Real life presents the opposite situation: a small number of discordant data which cannot be augmented with reasonable effort by new measurements. This situation is none of our imagination and characterizes not only research with B(illion) D(ollar) M(achines). It applies also to the quite different problem of the evaluation of physical constants and conversion factors from sophisticated high precision experiments carried out at different laboratories scattered all over the world. Let us illustrate by citation how the latter people proceed:'*After a thorough analysis using a number of least squares algorithms, the initial group of 38 items of stochastic input data was reduced to 22 by deleting those that were either highly inconsistent with the remaining data or had assigned uncertainties so large that they carried negligible weight.*' [1] In particular the second reason given for deleting data is curious to say the least. The selection of the 22 "good" data from the initial 38 was done in this case by CODATA, an interdisciplinary committee of the International Council of Scientific Unions. Scientific truth, however, is not a question of majority opinion. Half a page later in the same paper the authors comment on differences between the 1973 evaluation and the (still valid) 1986 results: '*The large change in K_V (a constant relating the SI unit of volt to a calibration standard) and hence in many other quantities between 1973 and 1986 would have been avoided if two determinations of F (the Faraday number) which seemed to be discrepant with the remaining data had not been deleted in the 1973 adjustment*' [1]. It is quite interesting that while deleting seemingly discrepant data in 1973 had turned out to be unacceptable, it has not prevented the committee to proceed in the same way in 1986.

The problem of outliers is so ubiquitous that it has also found its proper place in textbooks [2,3]. The proposed methods can more or less be characterized as testing the stability of a derived quantity against deletion of single data or groups of data in sophisticated but poorly justified manners. It is the purpose of this paper to reformulate and solve the problem in a consistent probabilistic framework. The present paper was stimulated by related ideas of Sivia [4] and Press [5].

2. Bayesian Analysis

For definiteness we consider in this section the problem of finding an appropriate arithmetic mean μ from a given set of data and associated experimental uncertainties. Our model equation reads

$$\mu = d_i + \xi_i \qquad , \tag{1}$$

where ξ_i are the experimental errors characterized by variances σ_i^2. Of course we assume that $\langle\sigma_i^2\rangle$ is testable information. According to the principle of maximum entropythe likelihood function for the model (1) becomes

$$p(\vec{d}|\mu, \vec{\sigma}, I) = \prod_i \frac{1}{\sigma_i\sqrt{2\pi}} \exp(-\frac{1}{2\sigma_i^2}(d_i - \mu)^2) \qquad , \tag{2}$$

where the product over i stems from the assumption of uncorrelated errors $\{d_i - \mu\}$. The quantity we need in order to calculate $\langle\mu\rangle$ and $\langle\Delta\mu^2\rangle$ from the experimental data $\{d_i, s_i\}$ is $p(\mu|\vec{d}, \vec{s}, I)$ which we obtain from Bayes theorem as

$$p(\mu|\vec{d}, \vec{s}, I) = \frac{p(\mu|\vec{s}, I)}{p(\vec{d}|\vec{s}, I)} p(\vec{d}|\mu, \vec{s}, I) \qquad . \tag{3}$$

Here s_i stands for the experimentally estimated value of σ_i. Let us assume no prior knowledge on μ. We code this in a flat improper prior. This is not harmful in the case of parameter estimationand since $p(\vec{d}|\vec{s}, I)$ is a normalizing constant in this context we have $p(\mu|\vec{d}, \vec{s}, I) \propto p(\vec{d}|\mu, \vec{s}, I)$. The problem has thus reduced to the calculation of the marginal likelihood $p(\vec{d}|\mu, \vec{s}, I)$. In order to calculate this probability from (2) we have to code our knowledge about $\vec{\sigma}$ in terms of $\vec{s}$. For completeness we first assume a set of good-natured data in which case the simplest assumption for the probability distribution of $\vec{\sigma}$, $p(\vec{\sigma}|\vec{s}, I)$ is

$$p(\vec{\sigma}|\vec{s}, I) = \prod_i \delta(\sigma_i - s_i) \tag{4}$$

from which we obtain by application of the sum rule

$$p(\vec{d}|\mu, \vec{s}, I) \propto \exp(-\frac{1}{2} \sum_i (d_i - \mu)^2 / s_i^2) \qquad . \tag{5}$$

This is a gaussian with average value $\overline{\mu}$ and variance Σ_μ^2 given by

$$\overline{\mu} = \Sigma_\mu^2 \sum_i d_i / s_i^2 \qquad , \; \Sigma_\mu^{-2} = \sum_i 1/s_i^2 \qquad . \tag{6}$$

This is the familiar weighted mean taught already in freshmen physics laboratory courses. If these formulae fail to produce acceptable results then it is quite obvious where to modify our assumptions. It may in particular be difficult to estimate experimental uncertainties to a precision such that (4) is justified. In such a case we would generalize (4) to

$$p(\vec{\sigma}|\vec{s}, \vec{\omega}, I) = \prod_i \delta(\sigma_i - s_i/\sqrt{\omega_i}) \qquad , \tag{7}$$

where we have introduced scale variables ω_i whose magnitude determine to what extent $\langle\sigma_i\rangle$ deviates from s_i. The particular parameterization with $\sqrt{\omega_i}$ has been chosen for later calculational convenience. Integrating out the true errors from (2) using (7) yields

$$p(\vec{d}|\mu, \vec{\sigma}, \vec{\omega}, I) \propto \prod_i \sqrt{\omega_i} \exp(-\frac{\omega_i}{2s_i^2}(d_i - \mu)^2) \qquad . \tag{8}$$

In order to get rid of the newly introduced parameter $\vec{\omega}$ we have to specify its probability distribution. This specification codes our belief of how well the experimentally determined $\vec{s}$ approximate $\vec{\sigma}$. We should certainly expect $\langle\sigma_i\rangle = s_i$ and therefore $\langle\omega_i\rangle = 1$, otherwise it means that the determination of $\vec{s}$ suffers from a known bias. In addition we consider $\Delta^2\omega$ as testable information. Along with the fact $\omega > 0$ an appropriate prior is the ?-distribution

$$?(\omega|a) = \frac{a^a}{?(a)}\omega^{a-1}\exp(-a\omega) \qquad , \tag{9}$$

for which $\langle\Delta\omega^2\rangle = 1/a$. The parameter a specifies therefore our knowledge about the width of $p(\omega|\langle\omega\rangle, a, I)$ and therefore somehow the confidence interval of the experimentally determined quantities $\{s_i\}$. Note that (9) approaches an exponential distribution for $a \to 1$ and Jeffreys' prior for $a \to 0$. We employ (9) to marginalize over ω in (8) and obtain

$$p(\vec{d}|\mu, \vec{s}, I) \propto \prod_i \left\{a + (d_i - \mu)^2/2s_i^2\right\}^{-(a+1/2)} \qquad , \tag{10}$$

from which we can now calculate $\langle\mu\rangle$ and $\langle\Delta\mu^2\rangle$. The decisive difference between (5) and (10) is obviously that (5) satisfies $\prod f(x_i) = f(\sum x_i)$ while (10) does not. The product of an arbitrary number of gaussians is again a gaussian and is strictly unimodal. (10) may well be multi-modal and will certainly be so in the case of discrepant data. For the special case $a = 1/2$ (10) reduces to a product of Lorentzians (Cauchy distributions). While the mean value μ obtained from (5) is an average weighted only by the individual $\{s_i\}$ regardless of the position of the data points, both error variance $\{s_i\}$ and position of $\{d_i\}$ enter the calculation of $\langle\mu\rangle$ and $\langle\Delta\mu^2\rangle$ from (10).

The analysis developed so far shall now be applied to two sets of precision measurements reported in [1]: The Quantum Hall(QHE) resistance as an example for a rather concordant set of data and the gyromagnetic ratio of the proton as a highly discrepant set of measurements. The former quantity has become the primary standard of electrical resistance with numerical value $h/e^2 = 25812.8056(12)\Omega$. h is Planck's constant and e the electron charge. The numbers in brackets indicate the uncertainty in the last digits. The second quantity $\gamma_p = 26752.2128(81)[10^4 s^{-1}T^{-1}]$ is the ratio of the protons magnetic moment divided by its angular momentum. By appropriate reformulation we find

$$\gamma_p = \frac{\mu_p}{I_p} = \frac{\mu_p}{\mu_n}\frac{F}{M_p} \qquad , \tag{11}$$

where μ_p and μ_n are the proton and the nuclear magnetic moment respectively, F is the Faraday number and M_p the proton molar mass. Regarding M_p as an available auxiliary constant the measurement of γ_p serves therefore primarily for a better determination of the Faraday number F. The left panel of fig. 1 shows the QHE data with their error bar. The dotted line is the distribution (5), that is the posterior probability for μ on the basis of weighted least squares. The horizontal scale variable is $\mu - \mu_0$ where μ_0 is the straightforward unweighted average of the data points. It has no significance other than an appropriate shift of the μ axis for graphical presentation of the data. The dashed line is the posterior probability for μ according to (10) for $a = 1/2$. Its maximum occurs at nearly the least squares value of μ. However, we observe that this posterior is slightly wider. This reflects the fact that the error in μ according to (10) depends unlike the case of (5) not only on the quoted error variances of the contributing measurements but also on the distribution of $\{d_i\}$ themselves. The right-hand panel of fig. 1 shows data on

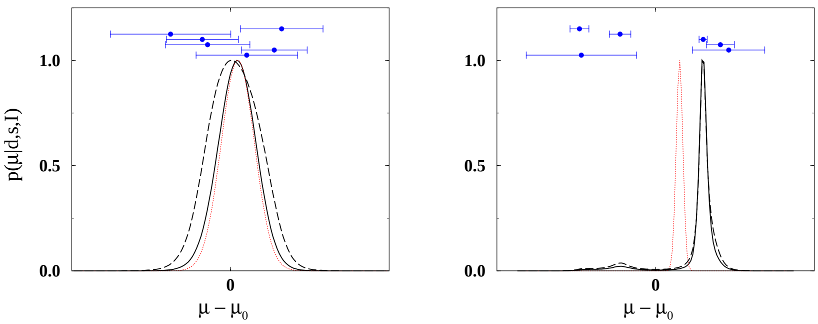

Figure 1. Posterior probability $p(\mu|\vec{d},\vec{s},I$ *for the QHE (left panel) and the* γ_p *(right panel) data. Dotted line (5), dashed line (10) and full curve (24)*

the proton gyromagnetic ratio(γ_p). This data set is of course very interesting. The dotted line indicates again the posterior probability for μ according to (5). From (6) we know already that its width is smaller than the error margin of the most precise of the contributing data. Together with the fact that its average lies in a region which does not overlap with any of the contributing data renders these results unacceptable. In fact the authors of [1] realized this fact themselves: '*Of all of the* γ_p *data, the most glaring discord comes from the NPL (UK) low field value (the leftmost data point in fig. 1). The measurements of the proton resonance frequency were completed in December 1975 after which the coil dimensions were measured, but no verification was made (by repeating the frequency measurement), that the measurement process did not affect the coils. Because this result is so discrepant, and because the measurements were forced to terminate prematurely we consider it to be an incomplete effort which should not be included in the final adjustment*'. Concerning deletion of data we refer the reader back to the introduction. The arbitrariness of deleting the most discrepant data point becomes quite obvious when we realize that even the truncated data set remains highly discrepant as also noted by the authors of [1]. While according to (6) they obtain a $\chi^2/N = 33$ for the full data set, the truncated set results in $\chi^2/N = 12.7$ instead of one.

The dashed curve in fig. 1 is the posterior probability for μ according to (10). It shows immediately the dilemma: The distribution is bimodal and a summary in terms of $\langle\mu\rangle$ and $\langle\Delta\mu^2\rangle$ is of very limited value. Note, however, that the prominent maximum lies in the range of the most precise of the contributing data points in accord with common sense. This result in not only numerically superior but it also points out that the data do not furnish a unique answer and that the data set might need a second thought. The straight forward application of the orthodox formula (5) could easily be deceiving.

For the present one-dimensional examples it is more or less intuitively clear upon display which data points should be regarded as an outlierand to what extent. This intuitive perception vanishes if we consider a multi-dimensional case that is a situation where μ represents a function of several variables. In particular in such cases it may be desirable to calculate $\langle\sigma_k\rangle$, the expectation value of the error variance of datum d_k in the light of the quoted error variance s_k and all the other data $\{d_i, s_i\}$. To this end we need $p(\vec{\sigma}|\vec{d}, \vec{s}, I)$ which is obtained from Bayes theorem as

$$p(\vec{\sigma}|\vec{d}, \vec{s}, I) = \frac{p(\vec{\sigma}|\vec{s}, I)}{p(\vec{d}|\vec{s}, I)} p(\vec{d}|\vec{\sigma}, I) \tag{12}$$

and

$$\langle\sigma_k\rangle = \frac{1}{p(\vec{d}|\vec{s}, I)} \int \sigma_k \, p(\vec{\sigma}|\vec{s}, I) p(\vec{d}|\vec{\sigma}, I) d\vec{\sigma} \quad . \tag{13}$$

$p(\vec{\sigma}|\vec{s}, I)$ is derived by application of the sum rule from (7) and (9) as

$$p(\vec{\sigma}|\vec{s}, I) = \int p(\vec{\omega}|I) p(\vec{\sigma}|\vec{s}, \vec{\omega}, I) d\vec{\omega} \quad . \tag{14}$$

Similarly we obtain $p(\vec{d}|\vec{\sigma}, I)$ from

$$p(\vec{d}|\vec{\sigma}, I) = \int p(\mu|I) p(\vec{d}|\mu, \vec{\sigma}, I) d\mu \quad . \tag{15}$$

Putting the various terms together we end up with

$$\langle\sigma_k\rangle / s_k = \frac{?\,(a)}{?\,(a + 1/2)} \int \left\{a + (d_k - \mu)^2 / 2 s_k^2\right\}^{1/2} \rho_a(\mu, d_i) d\mu / Z \quad , \tag{16}$$

with

$$Z = \int \rho_a(\mu, d_i) d\mu \quad \text{and} \quad \rho_a(\mu, d_i) = \prod_i (a + (d_i - \mu)^2 / 2 s_i^2)^{-(a+1/2)} \quad . \tag{17}$$

(16) allows to identify data point k as an outlierat a significance level δ if the right-hand side of (16) exceeds δ. The prior probability which we have used so far is (9) with $a = 1/2$. This represents the case where we assume that the expectation value of the true error variances σ_i is equal to the error s_i with a mean square deviation $\pm s_i\sqrt{2}$. It seems in order to exploit the influence of the choice of a on the final results. Fig. 2 shows such a scan of a together with mean value and posterior rms error again for the QHE data in the left panel and for the γ_p data in the right panel. Dashed curves apply in both cases to the above analysis. Since the smaller a the larger the range of the true error variance which we allow given the quoted variance, the behavior of the $\langle\Delta\mu^2\rangle^{1/2}$ is in accord with expectation. We would even accept a considerable variation of μ as a function of a in the case of discordant data as in the case of γ_p. The variation of $\langle\mu\rangle$ in the case of the QHE

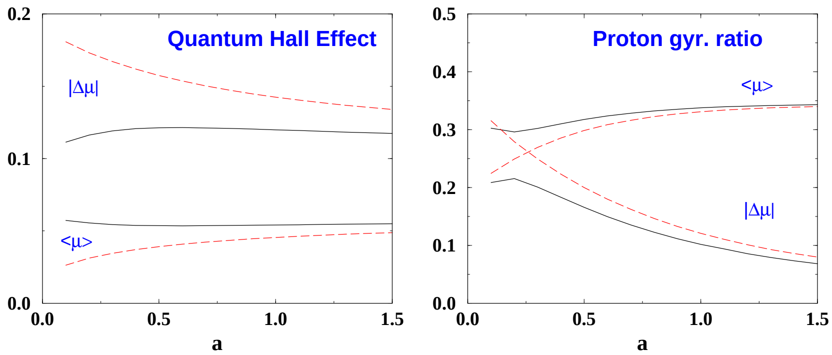

Figure 2. *Dependence of* $\langle\mu\rangle$ *and* $|\Delta\mu|$ *on the variance* $1/a$ *of (9)*

data is however less acceptable and calls for a modification of the analysis which was developed so far.

The assumption (7) that the quoted errors on every data point are uncertain according to the distribution (9) may be too strong. In fact, in real life we would expect that some of the quoted errors s_i should be treated with (4) while only a priori unknown outliers should be treated according to (7). Accordingly we modify the prior probability $p(\vec{\sigma}|\vec{s},\vec{\omega},I)$ introducing a new variable β with $0 \leq \beta \leq 1$ to

$$p(\vec{\sigma}|\vec{s},\vec{\omega},\beta,I) = \prod_i \{\beta\delta(\sigma_i - s_i/\sqrt{\omega_i}) + (1-\beta)\delta(\sigma_i - s_i)\} \qquad , \qquad (18)$$

where β is the probability that data point i is an outlierand the quoted error differs from the expectation value of the true error while $(1-\beta)$ is the probability that data point i is regular and the quoted error s_i is a good estimate for the expectation value $\langle\sigma_i\rangle$. For β we shall assume a flat prior in the interval [0,1]. The marginal probability $p(\vec{d}|\mu,\vec{s},I)$ is now obtained using (2), (18), (9) as

$$p(\vec{d}|\mu,\vec{s},a,I) \quad \propto \quad \int d\beta \prod_i \int d\omega_i \omega_i^{a-1} e^{-a\omega_i} \frac{a^a}{?(a)} \times$$
$$\int d\sigma_i \{\beta\delta(\sigma_i - s_i/\sqrt{\omega_i}) + (1-\beta)\delta(\sigma_i - s_i)\} \frac{1}{\sigma_i} e^{-\frac{1}{2\sigma_i^2}(d_i-\mu)^2} \qquad (19)$$

The σ_i integration is straightforward and results in a sum of two terms: the first has a structure similar to (8) multiplied by β and the second is a gaussian multiplied by $(1-\beta)$ with σ_i replaced by s_i. It is also straightforward to perform the ω_i integration. The second term in (19) does not contain ω_i and since $p(\omega_i|a)$ is normalized the ω_i integration results in a factor of 1. The second term is then according to σ_i and ω_i integrations

$$B_i = \exp[-(d_i-\mu)^2/2s_i^2]/s_i \qquad , \qquad (20)$$

the first term becomes

$$A_i = \frac{a^a ?\,(a+1/2)}{s_i ?\,(a)} \left\{ a + \frac{(d_i-\mu)^2}{2s_i^2} \right\}^{-(a+1/2)} \quad . \tag{21}$$

We may further suppress the common factor $1/s_i$ and are left with the β-integration

$$p(\vec{d}|\mu, \vec{s}, a, I) \propto \int d\beta \prod_i^N \{\beta A_i + (1-\beta) B_i\} \quad . \tag{22}$$

The product over the N factors is a polynomial of degree N in β

$$P_N = \sum_{i=0}^{N} \binom{N}{i} \beta^i (1-\beta)^{N-i} C_i^{(N)} \quad , \tag{23}$$

with so far unknown coefficients $C_i^{(N)}$. The integral in (23) over β of P_N in $0 \le \beta \le 1$ is the Beta function. After simplifying the binomial coefficients, (23) becomes

$$p(\vec{d}|\mu, \vec{s}, a, I) \propto \frac{1}{N+1} \sum_{i=0}^{N} C_i^{(N)}(\mu) \quad . \tag{24}$$

Our final task, the determination of the coefficients $C_i^{(N)}$, is accomplished by $P_N = P_{N-1}(A_N \beta + B_N(1-\beta))$, where we exploited the different representations (22) and (23) of P_N. Note that the unknown coefficients in P_N are $C_i^{(N)}$ while in P_{N-1} we have $C_i^{(N-1)}$. Inserting the explicit expressions (23) for the polynomials and equating terms with equal powers in $\beta^i(1-\beta)^j$ we obtain the recurrence relation

$$C_i^{(N)} = \frac{i}{N} A_N C_{i-1}^{(N-1)} + \frac{N-1}{N} B_N C_i^{(N-1)} \quad . \tag{25}$$

The recurrence is started from $P_1 = \beta A_1 + (1-\beta) B_1 = C_0^{(1)}(1-\beta) + C_1^{(1)}\beta$, which gives $B_1 = C_0^{(1)}$ and $A_1 = C_1^{(1)}$. This completes the calculation of the marginal likelihood $p(\vec{d}|\mu, \vec{s}, a, I)$. The solid solid line in fig. 1 and fig. 2 corresponds to the results of this last piece of analysis when applied to the two previous data sets. In the case of the QHE the posterior probability (24) falls close to the least squares posterior (5). The width of (24) is slightly larger than that of (5) which signals that the QHE data are nearly but not exactly compatible with each other. The posterior (10) reflects clearly the more diffuse prior for the exact errors in a noticeably larger width. An entirely different behavior is observed in the case of the γ_p data which are highly discordant. Here the posterior (24) coincides nearly with (10) while the least squares result differs grossly from the latter approaches. (24) has thus the extremely beneficial property that it interpolates between (5) and (10) depending on the condition of the data. This is further corroborated by fig. 2 which shows as solid lines $\langle\mu\rangle$ and $\sqrt{\langle\Delta\mu^2\rangle}$ for the QHE data in the left panel and for the γ_p data in the right panel. While the posterior $\langle\mu\rangle$ and $\langle\Delta\mu^2\rangle$

calculated from (24) do not longer depend on the particular choice of a in the distribution (9) in a wide range of a-values for the good-natured QHE data, there is only a minor difference in $\langle\mu\rangle$ and $\sqrt{\langle\Delta\mu^2\rangle}$ for the discordant γ_p data when calculated either from (24) or from (10). We conclude that (18) constitutes a very flexible prior distribution to code the frequent situation of unprecise knowledge about the errors associated with experimental data.

In the prior probability (18) we have introduced a new variable β. Its numerical value is the probability that a data point is an outlier. We are now in the position to calculate the posterior probability for β $p(\beta|\vec{d}, \vec{s}, a, I)$. From Bayes theorem, we have $p(\beta|\vec{d}, \vec{s}, a, I) \propto p(\vec{d}|\beta, \vec{s}, a, I)$ and using the sum rule

$$p(\beta|\vec{d}, \vec{s}, a, I) \propto \int p(\vec{d}|\mu, \beta, \vec{s}, a, I) d\mu \qquad . \tag{26}$$

The prior for μ, assumed flat, has been suppressed here. Formally (26) is equivalent to (19) with β-integration exchanged by a μ-integration. In terms of the polynomial coefficients $C_i^{(N)}$ the posterior probability for β becomes

$$p(\beta|\vec{d}, \vec{s}, a, I) \propto \int d\mu \sum_{i=0}^{N} \binom{N}{i} C_i^{(N)}(\mu)\beta^i(1-\beta)^{N-i} \qquad . \tag{27}$$

Fig. 3 shows the posterior probability distribution of β according to (27) for the two test data sets again for $a = 1/2$. For the QHE data the function (27) is

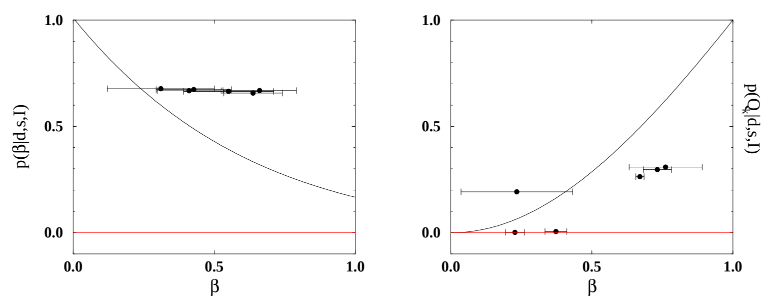

Figure 3. Average outlierprobability (continuous curves) and individual outlier probability for the QHE (left) and the γ_p data (right).

monotonous and peaks at $\beta = 0$ thus indicating a low average outlierprobability $\langle\beta\rangle$. For γ_p on the other hand the probability peaks at $\beta = 1$ and thus reflects that the γ_p data are highly discordant. The posterior probability for β answers therefore the interesting question as to what extent we deal with a good or a bad data set on the average.

We shall finally answer the question what is the probability that a single point d_k is either a good or a bad data point. Let us introduce the proposition Q_k: *d_k is* *an outlier*. We then need to calculate $P(Q_k|\vec{d},\vec{s},a,I)$. This probability is expressed by Bayes theorem as

$$P(Q_k|\vec{d},\vec{s},a,I) \propto \int P(Q_k|a,\beta,I)p(\vec{d}|Q_k,\vec{s},a,\beta,I)\,d\beta \quad , \tag{28}$$

where we have dropped the evidence denominator since it serves only as a normalization constant here. The marginal likelihood $p(\vec{d}|Q_k,\vec{s},a,\beta,I)$ is obtained from $p(\vec{d},\mu|Q_k,\vec{s},a,\beta,I)$ by marginalization over μ, again as before with a flat prior in μ. The prior probability $P(Q_k|a,\beta,I)$ is by definition just equal to β. The outlierprobability is finally

$$P(Q_k|\vec{d},\vec{s},a,I) \propto \int d\beta d\mu\ \beta\ A_k \prod_{i\neq k}(\beta A_i + (1-\beta)B_i) \quad . \tag{29}$$

Since we have abandoned normalization we must also calculate the probability for the proposition $\overline{Q_k}$: d_k *is not an outlier*. We obtain $P(\overline{Q_k}|\vec{d},\vec{s},a,I)$ from (29) by exchange of βA_k with $(1-\beta)B_k$. This completes our analysis for the generalized arithmetic mean in the case of partly discordant data. Fig. 3 shows that according to the above analysis the probability of the QHE data for being well conditioned is above 50% for all values. The γ_p data on the other hand fall all below 50%. Interestingly enough our analysis renders not only the NPL value entirely improbable but also the VNIIM (USSR) value which is at the position of the secondary maximum in fig. 1. The selective removal of the NPL value from the data evaluation process in [1] appears therefore even more doubtful.

3. Conclusions

Bayesian analysis has been applied to the frequent situation in experimental sciences that error bars on data can only inaccurately be specified. The analysis considers this imprecise knowledge and allows further for the possibility that data may originate from the assumed sampling distribution but that also outliers discordant with the rest of the data are present. The formulation in terms of "good" and "bad" data appears to be a sensible interpolation between the two limiting cases of entirely concordant and entirely discordant data. It should (and will) be put to further tests in more complicated cases of parameter estimation.

References

1. E. Cohen and B. Taylor *Reviews of Modern Physics*, **59**, p. 1121, 1987.
2. P. Rousseeuw and A. Leroy, *Robust Regression and Outlier Detection*, John Wiley and Sons, New York,Chichester, Brisbane, Toronto, Singapore, 1987.
3. J. Taylor, *An Introduction to Error Analysis*, University Science Books, Sausalito, California, 1997.
4. D. Sivia, "Dealing with duff data," in *MAXENT96 - Proceedings of the Maximum Entropy Conference 1996*, M. Sears, V. Nedeljkovic, N. E. Pendock, and S. Sibisi, eds., p. 131, NMB Printers, Port Elizabeth, South Africa, 1996.
5. W. Press in *Proceedings of Conference in Honor of John Bakcall*, J. Ostriker, ed., p. in press, Princeton University Press, Princeton, 1998.