

Reducing Subjectivity in the Likelihood

S. James Press

University of California, Riverside

Abstract: Some scientists prefer to exercise substantial judgment in formulating a likelihood function for their data. Others prefer to try to get the data to tell them which likelihood is most appropriate. We suggest here that one way to reduce the judgment component of the likelihood function is to adopt a mixture of potential likelihoods and let the data determine the weights on each likelihood. We distinguish several different types of subjectivity in the likelihood function and show with examples how these subjective elements may be given more equitable treatment.

Key words: Mixture likelihood, model averaging, subjectivity.

1. Introduction

We propose methods for modeling the likelihood function that will require fewer subjective judgments. We first discuss the nature of the problem of subjectivity in the likelihood function; then we review some related research; and finally, we define a mixture likelihood function and suggest estimation procedures that reduce the effects of subjective views imposed on the observed data.

1.1 Statement of the problem

It is sometimes desirable that beliefs of experimenters should be brought into a scientific analysis in ways that minimally distort the measured data (see, for example, Hogarth, 1980; Kyberg and Smokler, 1980; Lad, 1996). But that having been said, scientists observing data sometimes interpret the data points subjectively, according to what they want the data to show, and according to how precisely they believe the data points were measured. The latter procedure is of course quite common. This subjective interpretation of observed data may be totally at the unconscious level, or it may be purposeful (with the purposeful interpretation, the analysis may become fraudulent; see for example, Grayson, 1995, 1997; Howson and Urbach, 1990; and Press and Tanur, 2001).

The subjective interpretation of empirical data in medicine was discussed by Kaptchuk (2003). He stated (page 1, *op. cit.*):

Doctors are being encouraged to improve their critical appraisal skills to make better use of medical research. But when using these skills, it is important to remember that interpretation of data is inevitably subjective and can itself result in bias. Facts do not accumulate on the blank slates of researchers' minds, and data simply do not speak for themselves. Good science inevitably embodies a tension between the empiricism of concrete data and the rationalism of deeply held convictions. Unbiased interpretation of data is as important as performing rigorous experiments. This evaluative process is never totally objective or completely independent of scientists' convictions or theoretical apparatus.

Statistical analysis of a data set most often proceeds by summarizing the distribution of the data in terms of its likelihood function. In order to specify the form of the likelihood function, various assumptions are made about the data, such as mutual independence, identical distributions, unimodality, etc. After the likelihood function has been specified, additional assumptions are sometimes made (significance levels thought to be appropriate are specified, a prior distribution about the underlying unobservable quantities may be brought in, etc.). Analysis of the data generally proceeds by trying to keep the likelihood function treatment of the data as simple as possible, so that the scientist or analyst will introduce minimal distortion of the data. The analyst tries not to discard data, and tries to maximize the chance of understanding what nature is trying to tell us through the revealed data about the underlying phenomenon. In this way, when the analysis of the data has been completed, the claim can reasonably be made that the conclusions drawn from the analysis approximate, if not precisely reflect, the laws of nature, rather than the possible misinterpretations and misunderstandings of the laws of nature by human beings. It will be useful to first briefly define what we mean by objectivity and subjectivity, in this context.

According to Mandik (2001)¹,

The word *objectivity* refers to the view that the truth of a thing is independent from the observing subject. The notion of objectivity entails that certain things exist independently from the mind, or that they are at least in an external sphere. Objective truths are independent of human wishes and beliefs. The notion of objectivity is especially relevant to the status of our various ideas, and the question is to what extent objectivity is possible for thought, and to what extent it is necessary.

¹Mandik, P. (2001). *The Internet Encyclopedia of Philosophy*, see <http://www.utm.edu/research/iep/o/objectiv.htm>.

This is but one of many definitions that have been suggested. The elusive quest for objectivity in science has been, and remains, an important topic of discussion among historians and philosophers of science (for extensive additional discussions of the meaning of “objectivity”, see for example, Bower, 1998; Porter, 1995, 1996; and Daston and Galison 1992). For some, scientific objectivity involves the search for *certainty* in knowledge about one of nature’s well-kept secrets, independent of what human beings believe; but in many cases, we find that what we earlier thought to be true about nature, turns out later to be questionable.

In an interesting example from physics, Folger, 2003, pointed out that:

Pioneer 10, launched in 1972, is now some 8 billion miles from home. But it has been slowing down, as if the gravitational pull on it from the sun is growing progressively stronger the farther away it gets. Milgrom proposed (see the MOND pages-MODified Newtonian Dynamics)² that Newton’s laws might change at these accelerations. If Milgrom is right, Newton’s and Einstein’s laws will be in for some major tweaking.

Sometimes the scientist has such deep understanding and insight into the phenomenon he/she is studying that the scientist’s own predictions of what should be found from the analysis are far superior to what the data analysis seems to indicate. In some cases the beliefs of the scientist or analyst are so strong, even before actually taking any data that bear on the phenomenon, that the data are interpreted or manipulated so that they will reflect these preconceived views of the scientist. Any preconceived personal views (views held before taking any data), weak or strong, are what we refer to in this context as *subjectivity*.

1.2 Related Research

One approach to reducing the effects of differing assumptions about likelihoods may be found in a line of research that involves use of the *empirical likelihood function*. In this approach, most useful in large samples, a discretized, binned, version of the empirical cdf, instead of a specific likelihood function, is used. Inference is then made from a multinomial distribution. An unfortunate feature of this approach is the additional unknown parameters that are concomitantly introduced into the model. See: Owen, 1988, 2001. For typically small and moderate size samples this could be a problem, but for the massive data sets typical of data mining applications (see, for example: Berry and Linoff (1997); and Hastie, Tibshirani, and Friedman, 2001) such an approach could be a helpful alternative.

²MOND pages — <http://www.astro.umd.edu/ssm/mond/>

We show in the next section how we might understand and account for some types of subjectivity that sometimes enters the *likelihood function*, and might not be desired. We will use the definition and form of the likelihood function in which for absolutely continuous random variables, up to a proportionality constant, it is the joint probability density function of the observables given the unobservables.

2. Types of Subjectivity in the Likelihood Function

We distinguish three of the types of likelihood subjectivity problems that may occur:

- (a) how to determine the distributional form of the likelihood function in a way that is largely objective, but permits the data themselves to guide the modeling as to whether the data are Normally-distributed, or Gamma-distributed, or possibly follow some other convenient distribution. We call this problem, “*distributional subjectivity*”;
- (b) how to treat observed data that have possibly been weighted subjectively so that some data points are valued more heavily than others, and some are even ignored; we call this problem, “*weighted-data subjectivity*”;
- (c) how to account for the nature of the experiment used to obtain the data that may have favored one type of response over another; we call this problem, “*experiment subjectivity*”.

We treat each of these types of subjectivity in Section 3.

3. Reducing Likelihood Subjectivity

3.1 The mixture likelihood

We use a convex mixture of various likelihoods for the data; the usual likelihood function results as a special case.

Suppose an experiment is repeated n times with the resulting one-dimensional data outcomes: x_1, x_2, \dots, x_n . We suppose that there are J models for the data that potentially we might reasonably entertain. For simplicity, merely to suggest a general type of approach, we consider problems involving only one unknown parameter, namely, the means of the J distributions, ω .

In some situations, the parameters may be quite different from one another but they can generally be related functionally. For example, the case of distinguishing between the means of normal and log-normal distributions, where the mean parameter has different meanings in the two cases is sometimes particularly interesting. In such cases, functional relationships among the parameters are required.

Suppose, in the one-parameter problem, we can assume these data to be mutually independent and identically distributed, and we agree to adopt the likelihood function for Model m_j :

$$\ell_j(x_1, \dots, x_n | m_j, \omega) \equiv \ell_j(\underline{x} | m_j, \omega).$$

Define a “mixture likelihood function”,

$$L_M(x_1, \dots, x_n | \omega) \equiv L_M(\underline{x} | \omega),$$

such that:

$$L_M(\underline{x} | \omega) = E\{\text{likelihood}\} = E_{Model}[\ell(\underline{x} | \omega)] = \sum_{j=1}^J \ell_j(\underline{x} | \omega) P(m_j | \omega) \quad (3.1)$$

where $\ell_j(\underline{x} | m_j, \omega)$ denotes the usual likelihood function of the data under model m_j , $\ell(\underline{x} | \omega)$ denotes a model-independent likelihood function, and $P(m_j | \omega)$ denotes the prior probability of model m_j . The mixture likelihood function is of course a likelihood function itself. If there were only one model ($J = 1$), L_M reduces to the ordinary likelihood. The mixture likelihood function explicitly assumes that we should combine different models in a linear way. Other possibilities exist of course, and perhaps in certain cases, they are even more desirable. But because for a wide variety of cases, the linear assumption seems appropriate, we will retain this assumption throughout. We next address the issue of how to reduce model subjectivity (how to choose the weights).

3.2 Reducing “model subjectivity”

In some instances, the scientist has very strong, theory-based, beliefs about how the data were generated, and how the corresponding likelihood function should behave. In such instances, especially in small samples, the analyst should surely use that information to permit the desired likelihood function to emerge. In other situations where the scientist/analyst wants the data to speak as loudly as possible relative to the scientist’s pre-conceived beliefs, there is no unique way to accomplish this objective. The approach suggested here is to take equal weights in the mixture. Accordingly, take all $P(m_j | \omega)$ in eqn. (3.1) to be equal (discrete uniform distribution). This interpretation of equal treatment for the different models is:

(1) in keeping with the approach frequently used for weighting in mixture models to express indifference or ignorance among the various components in the mixture;

(2) it is the procedure suggested by Laplace when he adopted his Principle of Insufficient Reason (Laplace, 1812, 1814);

(3) it is consistent with a basic result of information theory that the distribution that corresponds to maximum entropy, or minimum information, is the uniform distribution.

This gives the mixture likelihood function (equally-weighted average likelihood):

$$L_M(\underline{x}|\omega) = \frac{1}{J} \sum_{j=1}^J \ell_j(\underline{x}|m_j, \omega). \quad (3.2)$$

For example, suppose there are just two potential models ($J = 2$) that might reasonably represent the data: $N(\omega, 1)$ and a Student t -distribution centered at ω , with 3 degrees of freedom (a fat-tail distribution that has a population mean). Then, the mixture likelihood function becomes:

$$L_M(\underline{x}|\omega) = \frac{1}{2} \left\{ \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x_i - \omega)^2\right\} + \prod_{i=1}^n \frac{m^{m/2}/B(1/2, m/2)}{[m + (x_i - \omega)^2]^{(m+1)/2}} \right\}. \quad (3.3)$$

Clearly each term in equation (3.3) is non-negative and integrates to one (with respect to \underline{x}), so $L_M\{\underline{x}|\omega\}$ is a bone fide likelihood function for the data (as would be the case whichever models we choose). In some situations, one scientist might favor the normal distribution for representing the distribution of the data, while another might favor the Student t -distribution. By using $L_M\{\underline{x}|\omega\}$ to represent the likelihood function for all inferences, the analyst reduces the model subjectivity in the description of the data distribution. Maximum likelihood estimation of ω is now more complicated numerically than it would be with use of either the normal or the Student t distributions separately, but the numerical problem is straightforward (see numerical example below) and easily generalizes to more than two possible ordinary likelihoods.

We next numerically illustrate the example suggested in this section of how to reduce model subjectivity when the models under consideration are the $N(\omega, 1)$ and the Student t_3 centered at ω . We randomly generated a total of 20 observations, 10 observations from t_3 , a Student t -distribution with 3 degrees of freedom centered at $x = 10$, and 10 observations from $N(10, 1)$. The resulting data are shown in columns 2 and 3 of Table 1a. Then, using the Newton-Raphson method, we calculated the mixed MLE. It is given at the bottom of Column 2 as: $\hat{\omega} = 9.9168$. To illustrate variability, there are four replications of this entire process shown in Table 1a; the four resulting mixed maximum likelihood estimates (mixed MLE's) are also shown in Table 1a.

Table 1a: Four replications of model subjectivity

	t	normal	t	normal	t	normal	t	normal
1	11.1861	9.6734	8.1266	11.182	8.9931	8.331	11.476	9.8325
2	9.9749	11.542	11.092	10.175	10.641	10.131	10.922	11.051
3	8.5632	10.259	11.374	11.720	9.8717	7.8108	10.270	10.642
4	5.5471	9.4442	9.3422	10.757	8.8824	8.3177	8.3906	9.0293
5	9.5188	10.779	13.189	9.8871	9.7579	9.4354	11.899	8.8359
6	10.994	9.3448	9.6007	9.715	10.385	10.092	9.5234	10.566
7	9.1875	9.9779	9.9069	9.7106	13.659	9.4326	14.559	9.495
8	11.283	9.2274	8.3785	9.8394	9.5543	10.361	10.177	10.247
9	8.7574	10.724	10.073	11.637	9.2285	9.0399	10.155	9.0938
10	7.9804	11.263	10.379	9.4837	11.274	9.7291	9.2795	10.366
Mixed MLE	9.9168		10.240		9.6237		10.116	

For comparison purposes, we also computed the separate ordinary MLE's assuming all 20 observations were generated from a normal, and then, that all 20 observations were generated from a t_3 distribution. Results are given in Table 1b.

Table 1b: Separate MLE's For normal and Student data

Normal MLE	9.761	10.278	9.7463	10.291
t_3 -MLE	9.924	10.187	9.6104	10.111

Thus, it may be seen that in the first instance, while the mixed MLE is 9.9168, the MLE assuming all 20 observations came from a normal is 9.7614, whereas the MLE assuming all 20 observations came from a t_3 is 9.9240. Results for the other 3 cases are shown in Tables 1a and 1b as well. Depending upon the assumptions made for the modeling, results for the mixture MLE obtained from the model averaging may differ substantially from those of the separate models, or not.

3.3 Reducing “weighted-data subjectivity”

We examine two distinct cases of weighted data subjectivity and model the two cases separately below.

Case 1 — Several Observers (Scientists) Rate the Same Data Points Differently

In this case, different observers (scientists) might interpret the same points differently. Some observers might view certain points as mistakes (outliers that were generated from different distributions from the other points), and therefore delete them from the analysis; and others might, according to their own beliefs, weight certain points more heavily than others (perhaps difficult-to-measure

points might be weighted less heavily because the error associated with the measurement might be greater than with most of the other points; perhaps certain points obtained were measured under censored conditions; etc.).

For simplicity, assume the data points are mutually independent. We define the likelihood function for Observer O_k as:

$$\ell(\underline{x}|\omega) = \prod_{j=1}^n [f(\delta_{jk}x_j | O_k, \omega)]^{p_k(\delta_{jk}x_j|O_k)}, \quad (3.4)$$

where: $p_k(\delta_{jk}x_j|O_k) = 1$, if Observer O_k includes the data point x_j in the analysis, and $p_k(\delta_{jk}x_j|O_k) = 0$ if not; δ_{jk} denotes the weight that Observer k places on observation x_j , $f(x_j|\omega)$ denotes the pdf (probability density function) of X_j , conditional on ω . The mixture likelihood function may be defined as:

$$\begin{aligned} L_M(\underline{x}|\omega) &= E\{\text{likelihood}\} = E_{data}\{\ell_k(\underline{x}|\omega)\} \\ &= \sum_{k=1}^K \ell_k(\underline{x}|\omega)P_k(O_k), \end{aligned} \quad (3.5)$$

where $P_k(O_k)$ denotes the prior probability that the data analyst places on the model that has been developed by Observer k . To be objective (or indifferent among the choices), in the sense we have been discussing, we take $P_k(O_k) = 1/K$, for all k . Then,

$$L_M(\underline{x}|\omega) = \frac{1}{K}\ell_k(\underline{x}|\omega) \quad (3.6)$$

As a simple example, suppose that all K observers adopt the same distribution for the data, say, $N(\omega, 1)$ (in Section 3.2 the analyst adopted two different possible distributions for the data), and assume that they weight the points in the same way, so that $\delta_{jk} = 1$ for all k , for all points they include in their analyses, but they may include different points. Then, since the n observations are independent,

$$\ell_k(\underline{x}|\omega) = \prod_{j=1}^n \left[\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x_j - \omega)^2\right\} \right]^{p_k(x_j|O_k)}. \quad (3.7)$$

To be specific, suppose that $n = 102$, and that there are two observers, O_1 and O_2 . Suppose further that O_1 believes x_{102} is an outlier, and O_2 believes that both x_{101} and x_{102} are outliers, but they agree that the first 100 points (x_1, \dots, x_{100}) should be included in their analyses. Then,

$$\begin{aligned} p_1(x_j|O_1) &= 1 \quad \text{for } j = 1, 2, \dots, 101, \\ &= 0, \quad \text{for } j = 102, \end{aligned}$$

also,

$$\begin{aligned} p_1(x_j | O_2) &= 1 \quad \text{for } j = 1, 2, \dots, 100, \\ &= 0, \quad \text{for } j = 101, 102, \end{aligned}$$

Then,

$$\ell_1(\underline{x} | \omega) \equiv \ell_1 = \left(\frac{1}{\sqrt{2\pi}} \right)^{101} \exp\left\{-\frac{1}{2} \sum_{j=1}^{101} (x_j - \omega)^2\right\}, \quad (3.8)$$

and

$$\ell_2(\underline{x} | \omega) \equiv \ell_2 = \left(\frac{1}{\sqrt{2\pi}} \right)^{100} \exp\left\{-\frac{1}{2} \sum_{j=1}^{100} (x_j - \omega)^2\right\}, \quad (3.9)$$

Then,

$$\begin{aligned} L_M(\underline{x} | \omega) &= \frac{1}{2} \left\{ \left(\frac{1}{\sqrt{2\pi}} \right)^{101} \exp\left\{-\frac{1}{2} \sum_{j=1}^{101} (x_j - \omega)^2\right\} \right. \\ &\quad \left. + \left(\frac{1}{\sqrt{2\pi}} \right)^{100} \exp\left\{-\frac{1}{2} \sum_{j=1}^{100} (x_j - \omega)^2\right\} \right\} \quad (3.10) \end{aligned}$$

We may now estimate ω by maximizing $L_M(\underline{x} | \omega)$ with respect to ω . Note first that if we let n_1, n_2 be the numbers of data points used in the respective analyses of Observers O_1 and O_2 , they are also the numbers of terms in the two summations, and in this example, $n_1 = 101$ and $n_2 = 100$. We may readily find by ordinary differentiation, the mixture maximum likelihood estimator (mixture MLE) to be:

$$\omega = \alpha(\omega)\bar{x}_1 + [1 - \alpha(\omega)]\bar{x}_2, \quad 0 \leq \alpha(\omega) \leq 1, \quad (3.11)$$

where:

$$\alpha(\omega) \equiv \frac{n_1 \ell_1(\omega)}{n_1 \ell_1(\omega) + n_2 \ell_2(\omega)} \quad (3.12)$$

$$\bar{x}_1 \equiv \frac{1}{101} \sum_{j=1}^{101} x_j, \quad \bar{x}_2 \equiv \frac{1}{100} \sum_{j=1}^{100} x_j. \quad (3.13)$$

That is, we find the interesting result that $(\hat{\omega} | \ell_1, \ell_2)$ is a weighted average (actually a convex combination) of the separate MLE's that the two observers might adopt separately, and the weights are their respective proportions of their ordinary likelihoods, an intuitively sensible result. But note that because $\alpha(\omega)$ depends upon ω , equations (3.11) and (3.12) must be jointly solved numerically for $\hat{\omega}$.

While in large samples, the (continuous) data will generally ultimately swamp any prior distribution weights placed on the data points (see Le Cam, 1956), in small or moderate size samples, certain very influential points that may have been deleted from an analysis can have substantial effects on the interpretation of the experiment outcomes.

We next illustrate this example numerically. We randomly generated 18 points from $N(0, 1)$. We then ordered the points, and added 2 larger outliers. We assumed the first observer dropped the largest point as an outlier, and the second observer dropped the two largest points as outliers. We then calculated the mixture MLE numerically from equation (3.11) using the Newton-Raphson method. We replicated the procedure four times to examine variation. Data are shown in Table 2a.

Table 2a: Four replications of weighted-data subjectivity

Observation	$N(10, 1)$	$N(10, 1)$	$N(10, 1)$	$N(10, 1)$
1	8.3959	7.6748	8.1260	7.7977
2	8.4063	7.8796	8.5249	8.8122
3	8.559	7.9954	9.6225	8.9922
4	8.7975	8.7684	9.6490	9.0079
5	9.3082	8.9002	9.7041	9.0501
6	9.6001	8.9819	9.7444	9.1783
7	9.8433	9.2957	9.7660	9.2580
8	9.9802	9.3553	10.0400	9.3645
9	10.2570	9.3687	10.1180	9.4404
10	10.5710	9.5069	10.3150	9.7344
11	10.6690	9.6790	10.4280	9.8685
12	10.6900	9.8179	10.5690	10.0880
13	10.7120	9.8868	10.5780	10.2120
14	10.7140	10.0860	10.6230	10.2380
15	10.8160	10.3790	10.6770	10.3900
16	10.8580	10.4620	10.7310	10.4440
17	11.1910	10.5510	10.7990	10.5690
18	11.2540	10.9440	10.8960	10.7810
19	12.0000	12.0000	12.0000	12.0000
20	13.0000	13.0000	13.0000	13.0000

Calculations of MLE's for the data in Table 2a are given in Table 2b:

Table 2b: MLE's for data with outliers

Mixture MLE	10.0406	9.4205	10.0568	9.6267
\bar{x}_{18}	10.0346	9.4185	10.0506	9.6237
\bar{x}_{19}	10.1380	9.5544	10.1532	9.7487

We see that for the data in column 2 of Table 2b, for example, the mixture MLE was 10.0406. Had the observers carried out separate MLE's, with Observer 1 dropping only the last observation, he would have found his MLE to be 10.1380, while Observer 2 who dropped both of the last 2 observations would have found her MLE to be 10.0346. While the differences are not large they are intended to be illustrative.

Case 2 — One Observer (Scientist) Rates Each Data Point Differently

The second case of weighted-data subjectivity involves a single scientist weighting the importance of the data points differently from one another. Here we envision a single scientist who has carried out an experiment many times, but sometimes, for one reason or another, the scientist carried out the experiment with extremely small error, whereas on some other occasions, the scientist associated the experimental outcomes with considerably more error. Thus, which observed results had small associated error, and which had large associated error might differ from one replication of the experiment to the next.

In this context there is just one scientist who rates his/her experimental data differentially, according to how "well" the data point was measured, or what he/she thought should have occurred, or whatever. This is the more typical situation, compared with the first case. The mixture likelihood function is obtained from equations (3.4) and (3.5), for $K = 1$, as:

$$L_M(\underline{x} | \omega) = \ell_1(\underline{x} | \omega) P_1\{O_1\} = \prod_{j=1}^n [f(\delta_{j1} x_j | O_1, \omega)]^{p_1(\delta_{j1} | O_1)} . \quad (3.14)$$

To follow the paradigm suggested here we should take $\delta_{j1} = 1$ for every j . Of course the individual scientist would often argue that he/she knows better than anyone else that certain points were really not as good as others, and should therefore be down-weighted.

A now-classical example of this type of subjectivity of special historical interest has been documented with real data. It involves the data collected by R. A. Millikan (1868-1953). Dr. Millikan was an American physicist who successfully measured the charge on a single electron, winning a Nobel Prize in 1923 for this famous oil-drop experiment (as well as other prizes). Holton (1978) scrutinized Millikan's laboratory notebooks and found that Millikan had repeated his

oil-drop experiment 39 times, obtaining outcomes: x_1, \dots, x_{39} for the charge on the electron. Holton reported that Millikan had given each of his original sets of observations a personal quality-of-measurement rating: “best”, “very good”, “good”, “fair”, and no rating at all for discarded measurements (we interpret his weights to represent his prior probabilities for these measurements). The distribution of his rating results is summarized in the Table 3.

Table 3: Millikan’s measurements

rating descriptions	effective raating	$\delta_{j1} = \text{Weight}$	number of measurements
best	4	4/10	2
very good	3	3/10	7
good	2	2/10	10
fair	1	1/10	13
discard	no rating	—	7

For Millikan, $p_1(\cdot) = 1$, for 32 data points and $p_1(\cdot) = 0$ for the discarded 7 points. We order the measurements according to their effective ratings, from “best” to “fair”, and form the weighted average. The estimated value of the charge on the electron is then given by the weighted average:

$$\hat{e} = \frac{4}{10} \sum_{j=1}^2 x_j + \frac{3}{10} \sum_{j=3}^9 x_j + \frac{2}{10} \sum_{j=10}^{19} x_j + \frac{1}{10} \sum_{j=20}^{32} x_j.$$

Millikan formed the weighted average of his measurements and accordingly estimated the charge on the electron as 4.85×10^{-10} esu (electrostatic units). The ordinary equally weighted average would have been 4.70×10^{-10} esu. In his reported value he also averaged in the values obtained by other researchers. By contrast, the accepted value for “ e ”, the charge on the electron, today, is 4.77×10^{-10} esu. But the impressive closeness of Millikan’s values with today’s accepted value is deceptive; it occurred only because his values were based upon, “a faulty value for the viscosity of air, which when corrected, increases the discrepancy with the modern value by over 40%” (Mathews, 1998).

3.4 Reducing “experiment subjectivity”

Suppose there are two experiments that might be performed: E_g (“ g ” for “good”), and $E_{\bar{g}}$ (“ \bar{g} ” for “not good”). In E_g the scientist knows that the experiment will contain one or more variables that might produce effects that will be confounded with the effect of fundamental interest. In E_g , there are likely

to be fewer such confounding variables, so the scientist believes that he/she is more likely to be able to distinguish the effect he/she is seeking. Concomitantly, it may be that by carrying out E_g , the scientist is missing the important variables that suggest that the effect sought is really artifactual, and the seeming effect is explainable in other ways. Because the scientist is so convinced that the effect sought is real and not artifactual, he/she reasons that E_g is a “cleaner” and more promising experiment. The scientist might even argue, in a moment of enthusiastic zeal, that E_g is cheaper and/or less subject to error.

In both experiments, for simplicity of interpretation, we assume the data are normally distributed with variance equal to 1. Suppose that the scientist referred to above, call him/her Scientist A, would like to show that the population mean for the underlying phenomenon of interest is positive. If Scientist A carries out E_g , it is more likely that the sample mean \bar{x} will be positive than if Scientist A carries out $E_{\bar{g}}$ wherein the sample mean \bar{y} will imply the alternative hypothesis $H_{\bar{g}}$: that the population mean is not positive. If $E_{\bar{g}}$ is performed the scientist believes results are either unlikely to be supportive of the theory, or they are likely to be sufficiently marginal so that the theory will be in doubt. A priori, the experimenter adjudges the chances for concluding H_g : the population mean is positive, when performing E_g as greater than the chances for concluding that the population mean is positive when performing $E_{\bar{g}}$. Consequently, Scientist A decides to perform E_g .

Suppose some other scientist, say Scientist B, performs $E_{\bar{g}}$, and subsequently observes \bar{y} (using the same sample size, n). Let θ denote an indexing parameter such that $\theta = 1$ if the hypothesis H_g is true, and $\theta = 0$ if the hypothesis $H_{\bar{g}}$ is false.

$$\begin{aligned} L_M\{\text{data} | \theta\} &= E\{\text{likelihood}\} = E_{\text{experiment}}[\ell(\text{data} | \theta)] \\ &= \ell(\bar{x} | E_g, \theta)P\{E_g\} + \ell(\bar{y} | E_{\bar{g}}, \theta)P\{E_{\bar{g}}\} \end{aligned}$$

The mixture likelihood function becomes:

$$L_M\{\text{data} | \theta\} = P\{E_g\} \frac{\sqrt{n}}{\sqrt{2\pi}} \exp\left\{-\frac{n}{2}(\bar{x} - \theta)^2\right\} + P\{E_{\bar{g}}\} \frac{\sqrt{n}}{\sqrt{2\pi}} \exp\left\{-\frac{n}{2}(\bar{y} - \theta)^2\right\}.$$

An investigator cognizant of both experiments has both and available. In the same spirit of a desire for equity of treatment in the likelihood function, the investigator takes $P\{E_g\} = P\{E_{\bar{g}}\} = 0.5$. Then,

$$L_M\{\bar{x}, \bar{y} | \theta\} = \frac{1}{2} \left[\frac{\sqrt{n}}{\sqrt{2\pi}} \exp\left\{-\frac{n}{2}(\bar{x} - \theta)^2\right\} + \frac{\sqrt{n}}{\sqrt{2\pi}} \exp\left\{-\frac{n}{2}(\bar{y} - \theta)^2\right\} \right].$$

Define $z = (\bar{x} + \bar{y})/2$. Then, combining terms shows that:

$$L_M\{z | \theta\} = \frac{1}{2} \frac{\sqrt{n}}{\sqrt{2\pi}} \exp\{-n/4\} \exp\{-n(z - \theta)^2\}.$$

Thus, the MLE for θ is clearly: $\hat{\theta} = z = (\bar{x} + \bar{y})/2$. If Scientist A were correct in his/her a priori assessments of what was likely to happen in the experiment, $\hat{\theta}$ is likely to be closer to zero than \bar{x} (or even negative), a result that would tend to vitiate Scientist A's conclusions.

For example, for Scientist A's experiment, E_g , we generated 100 observations from $N(1, 1)$ and found $\bar{x} = 1.0598$. Then, for Scientist B's experiment, $E_{\bar{g}}$, we generated 100 observations from $N(-1, 1)$ and found $\bar{y} = -.9531$. So the generalized MLE, $\hat{\theta}$, is 0.053, a sample value just barely positive, which might not be convincing in many contexts for asserting that the population mean is really positive.

4. Conclusions

We have been concerned with how to reduce the effects of a scientist's pre-conceived beliefs in the analysis of his/her supposedly objectively-observed data. We have found that we can reduce the effect of some of those subjective interpretations by using a mixture likelihood function, and then choosing the mixture weights that weigh the various interpretations of the data equally.

References

- Berry, M. J., and Linoff, G. (1997). *Data Mining Techniques*. John Wiley.
- Bower, B. (1998). Objective visions: Historians track the rise and times of scientific objectivity. *Science News* **154**, 360-362.
- Daston, L. J. and Galison, P. (1992). The image of objectivity. *Representations* **40**, 81-128.
- Folger, T. (2003). Nailing down gravity: New ideas about the most mysterious power in the universe. *Discover Magazine*, Oct., 2003, 34-41.
- Grayson, L. (1995). *Scientific Deception*. The British Library.
- Grayson, L. (1997). *Scientific Deception – An Update.*: The British Library.
- Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning*. Springer-Verlag.
- Hogarth, R. (1980). *Judgment and Choice*. John Wiley.
- Holton, G. (1978). Sub-electrons, presuppositions, and the Millikan-Ehrenhaft dispute. In *The Scientific Imagination: Case Studies* (Edited by Gerald Holton), 25-83. The Cambridge University Press.
- Howson, C. and Urbach, P. (1990). *Scientific Reasoning: The Bayesian Approach*. Open Court Publishing Co.

-
- Kaptchuk, T. J. (2003). Effect of interpretive bias on research evidence. *British Medical Journal* **326**, 1453-1455.
- Kyberg, H. E. Jr., and Smokler, H. E., Editors (1980). *Studies in Subjective Probability*. Robert E. Krieger Publishing Co.
- Lad, F. (1996). *Operational Subjective Statistical Methods: A Mathematical, Philosophical, and Historical Introduction*. John Wiley.
- Laplace, P. S. (1812). *Theorie Analytique des Probabilities*.³ Paris: Courcier.
- Laplace, P. S. (1814). *Essai Philosophique sur les Probabilities*.⁴ Paris.
- Le Cam, L. (1956). On the asymptotic theory of estimation and testing hypotheses. *Proc. Third Berkeley Symposium on Math. Statist. And Prob.* **1**, 128-156. University of California Press. p. 308. ***** Please check where the p.308 come from? ***
- Mathews, Robert A. J. (1998). Facts versus factions: The use and abuse of subjectivity in scientific research. Cambridge, England: The European Science and Environment Forum, Working Paper 2/98, September, 1998.⁵
- Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237-249.
- Owen, A. (2001). *Empirical Likelihood*. Chapman and Hall.
- Porter, T. (1995). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton University Press.
- Porter, T. (1996). Statistics, social science, and the culture of objectivity. *Oesterreichische Zeitschrift fur Geschichtswissenschaften* **7**, 177-191.
- Press, S. J., and Tanur, J. M. (2001). *The Subjectivity of Scientists and the Bayesian Approach*. John Wiley and Sons.

Received January 12, 2005; accepted June 15, 2005.

S. James Press
Distinguished Professor Emeritus
Department of Statistics
University of California, Riverside
Riverside, CA, USA
jpress@ucr.edu

³The second, third, and fourth editions appeared in 1814, 1818, and 1820, respectively. It is reprinted in *Oeuvres Completes de Laplace*, Vol. VII, 1847, Paris: Gauthier Villars.

⁴This book went through five editions (the fifth was in 1825) revised by Laplace. The sixth edition appeared in English translation by Dover Publications, New York, in 1951. While this philosophical essay appeared separately in 1814, it also appeared as a preface to his earlier work, *Theorie Analytique des Probabilities*.

⁵See <http://www.esef.org>, p.5.