Robustizing Model Fitting

January 20, 1996 at 19:10

Michael P. Windham

Mathematics & Statistics
Utah State University
Logan, UT 84322–3900 USA

For your edification, entertainment, information ...
Comments sincerely requested

Abstract

A robust method for fitting a model from a parametric family is fundamental to effective statistical analysis. A procedure is given for robustizing any model fitting process using weights from the family from which the model is to be chosen. The weighting reduces the influence of information not compatible with the model family, while maintaining the basic structure of a familiar model fitting process. The procedure produces a parametric family of model fitting functions. The value of the parameter determines the degree to which the weighting influences the robustized model fit. A mechanism for determining an appropriate value for the parameter is described.

Keywords: ROBUST PARAMETER ESTIMATION; M-ESTIMATION; ITERATIVE REWEIGHTING; MAXIMUM LIKELIHOOD; METHOD OF MOMENTS

1. INTRODUCTION

Model fitting is one of the basic activities in statistical analysis. Parameter estimation fits a model from a parametric family, but the most common parameter estimation procedures, such as maximum likelihood or method of moments, are not robust. They are sensitive to outliers or other contamination in data, or perhaps to incorrect assumptions about an appropriate population model.

Since robustness requires accommodating deviations from a model, it would seem reasonable that the structure of the model family play a direct role in obtaining the fit. In particular, for a given model, suppose one could weight data so that data likely to occur under that model had relatively large weights and data not likely to occur had relatively small weights. One could then apply a parameter estimation method to obtain estimates for meaningful quantities, such as moments, or maximum likelihood estimates for the *weighted* data. The analogous quantities could be calculated for the model itself, that is for the distribution obtained by using the model to weight *itself*. If the estimate for the weighted data is close to its theoretical counterpart for the model , then one might believe that much of the data are likely to occur in a sample from the model. Therefore, the most reasonable fit of this kind would be the model for which the weighted data estimates and the theoretical values are the same. The method presented here finds this model.

The weights in the procedure I am proposing are proportional to a power of the density. Suppose, for example, that one wants to fit a normal with mean zero to data $\{x_1, ..., x_n\}$. The densities in the appropriate normal family are $\phi(x;\theta)$ where θ is the unknown variance. For a particular θ, we weight the j-th data point with $w_j = K\phi^c(x_j;\theta)$, where c is a fixed positive exponent and K is chosen so that the the weights sum to one. The exponent can be used to control the influence the model has on the weights. The larger the exponent the more data compatible with the model are weighted relative to those data not compatible. The variance of the weighted data would then be $v = \sum_j w_j x_j^2$. This value would be compared to $\tau = \int x^2 w(x)\phi(x;\theta)\, dx$, where $w(x) = K^*\phi^c(x;\theta)$. The constant K^* is chosen so that $\int w(x)\phi(x;\theta)\, dx$ is one, that is, so that $w\phi = K^*\phi^{c+1}$ is a density. This density is, in fact, normal with mean zero and variance $\tau = \theta/(c+1)$. If the normal model is a good fit to the bulk of the data, then τ and v should be close to each other. The method I will describe is an iterative procedure to determine the model that makes them equal.

The exponent c can be used as a tuning constant. The greater the exponent, the more the structure of the model family influences the fit. I will show in Section 4, that there is an analytical basis for choosing the exponent automatically. In particular, I will define a criterion $\rho(c)$ that measures the efficiency in using a particular member of the model family to describe a sample using weighting. The maximum of ρ occurs for an "optimal" choice of c. Most, if not all, robust parameter estimation methods have tuning constants, but they are either chosen by the user or given default values suitable for specific situations which may or may not be applicable. The method introduced here has the advantage that no such user choices are required.

I will refer to the method as *robust model fitting.*

2. EXAMPLES

Robust model fitting can be implemented in a variety of situations. As an illustration I will describe how it is applied to the most familiar, maximum likelihood estimation for normal models. Figure 1.a shows a histogram of a sample of size 200 from a standard univariate normal $(\mu = 0, \sigma^2 = 1)$ to which 30 data values near -7.0 have been added. The usual maximum likelihood estimates for the parameters are the sample mean, $\hat{\mu} = -.98$, and sample variance, $\hat{\sigma}^2 = 7.35$. The normal density with these parameters is also plotted with the histogram and is an understandably poor fit. Robustizing is accomplished by an iterative procedure that successively recomputes parameter estimates in two steps. The starting values are the maximum likelihood estimates for the unweighted data.

The first step in an iteration is to weight data with a power of the model density using the current parameter estimates, $\hat{\mu}$ and $\hat{\sigma}^2$. In this example, the weight for a data point x is proportional to $\phi^c(x; \hat{\mu}, \hat{\sigma}^2)$, where for the purposes of illustration, the exponent c was chosen to be .5. A histogram of the weighted data in the first iteration is shown in Figure 1.b. The effect of the weighting is emphasized by the horizontal lines in Figures 1.a and b. Data compatible with the model are given increased weight, while data not compatible are down weighted. The maximum likelihood estimates for the weighted data, call them $\hat{m}$ and $\hat{s}^2$, are then computed to be -.33 and 2.57, respectively. The corresponding normal density is also shown in Figure 1.b.

The second step in an iteration is to obtain the parameters for the model that would produce $\hat{m}$ and $\hat{s}^2$, if maximum likelihood estimation were applied to the model weighted by itself. A normal distribution with parameters μ and σ^2 weighted by the c-th power of itself is also a normal distribution, that is, $\phi^c(x; \mu, \sigma^2)\phi(x; \mu, \sigma^2)$ is proportional to $\phi(x; \mu, \sigma^2/(c + 1))$. For $c = .5$, the model with parameters $\hat{\mu}_+ = \hat{m} = -.33$ and $\hat{\sigma}^2_+ = 1.5\hat{s}^2 = 3.85$ when weighted by itself is a normal with parameters $\hat{m}$ and $\hat{s}^2$. The normal density with these parameters is shown in Figure 1.c, (solid line) super imposed on the histogram of the unweighted data. The original density is also plotted in the figure (dashed line) to show the improvement in fit from one iteration.

The iterations are continued, replacing $\hat{\mu}$ and $\hat{\sigma}^2$ with $\hat{\mu}_+$ and $\hat{\sigma}^2_+$, until convergence, that is, until the change in the parameters from one iteration to the next is sufficiently small. The table in Figure 1 shows the estimates for succeeding iterations. Figure 1.d shows the model obtained at convergence, which has parameters, $\hat{\mu} = -.04$ and $\hat{\sigma}^2 = 1.01$. It is this model that produces the same result when used to weight the data and when used to weight itself.

The optimal exponent based on the criterion ρ (Section 4) is $c = .492$ and the corresponding parameter estimates are $\hat{\mu} = -.04$ and $\hat{\sigma}^2 = 1.01$ or $\hat{\sigma} = 1.00$.

Since the parameters in this example are location and scale parameters, the estimates can be compared to other standard robust location and scale estimators. Location was estimated by the median and M-estimation using the Huber and bisquare ψ functions. Scale was estimated using the median absolute deviation (MAD), a τ-estimator using the Huber ψ function, and an A-estimator using the bisquare ψ function. The estimates were obtained in S^+ using defaults for all user choices. The defaults are chosen so that the estimators are consistent for estimation from standard univariate normal samples, and so, are appropriate for this example. The reader is referred to the documentation and program listings in S^+ for details. For location, the median is $-.26$, the Huber estimate is $-.29$, and the bisquare is $-.04$. For scale, the median absolute deviation (MAD) is 1.19, the Huber-τ is 1.30, and an A-estimate is 1.08. For this example, the *robustized model fit* is at least as good as these other robust estimators.

There are many robust estimators for the normal family. In a sense, most robust estimators are built with the normal family in mind, that is, they are most effective when uncontaminated data are expected to be symmetrically distributed about a central location. Dealing with asymmetric data is more difficult. Robust model fitting requires only a reasonable model for the asymmetry. For example, Figure 2 shows the histogram for a sample of size 400 from a gamma distribution with shape parameter $\alpha = 5$ and scale $\beta = 2$, to which 20 points near 35 have been added. The usual method of moments estimators for the parameters are $\hat{\alpha} = 2.8$ and $\hat{\beta} = 4.1$. The dashed line in Figure 2 shows the density with these parameters. The solid line is the robust model, with parameters 5.5 and 1.9, obtained with exponent $c = 1$. This model was was obtained by finding parameter estimates and values of the criterion ρ for exponents c determined by the golden section numerical optimization routine. This routine systematically searches for a maximum of ρ (Press et al, 1990, p. 293).

For a fixed value of c, parameter estimates were found by iterating the two step process as follows. The function $g(x; \alpha, \beta)$ is a gamma density. For current values $\hat{\alpha}$ and $\hat{\beta}$,

1. obtain weighted moments

$$\begin{aligned} m &= \sum_j w_j x_j \\ v &= \sum_j w_j (x_j - m)^2 \end{aligned}$$

   where w_j is proportional to $g^c(x_j; \hat{\alpha}, \hat{\beta})$.

2. New estimates are

$$\hat{\alpha}_+ = (c+1)\frac{v}{m} \quad (1)$$

$$\hat{\beta}_+ = \left(\frac{m^2}{v} + c\right) / (c+1) \quad (2)$$

The new estimates are the parameters of a gamma distribution which when weighted by itself is a distribution with mean m and variance v. In fact, a gamma with parameters α and β, weighted by itself is a gamma with parameters $c(\alpha - 1) + \alpha$ and $\beta/(c+1)$, so, (1) and (2) are just the method of moments calculations for the weighted gamma. The table in Figure 2 gives the parameter estimates and criterion value for a range of values of c.

3. GENERAL DESCRIPTION

I will adopt the notation commonly used in robust statistics to give a general description of the robustizing process. Assume a parametric family of densities $\{g(x; \theta)\}$ with θ in a subset Θ of R^k. A parameter estimator is a functional $T : \mathcal{F} \to \Theta$, where $\mathcal{F}$ is a general family of probability distributions. A data set is represented by its empirical distribution, that is the parameter estimate for a data set would be denoted $T(F)$, where F is the empirical distribution of the data.

For a given $c \geq 0$ and t in Θ, the weighting process can be formulated in terms of a weighted distribution. Namely, for F in $\mathcal{F}$, let $F_{c,t}$ be defined by $dF_{c,t}(x) = w(x; t)dF(x)$ with $w(x; t) = g^c(x; t)/E_F[g^c(X; t)]$. A robustization

TABLE 1: The function τ_c for parameter estimation using method of moments.

G_θ	θ	$(G_\theta)_{c,\theta}$	$\tau_c(\theta)$
Normal	μ, Σ	Normal	μ, $\Sigma/(c+1)$
Exponential	λ	Exponential	$\lambda/(c+1)$
Gamma	α, β	Gamma	$(c+1)\alpha - c$, $\beta/(c+1)$
Beta	α, β	Beta	$(c+1)\alpha - c$, $(c+1)\beta - c$
χ^2	ν	Gamma	$\nu - 2c/(c+1)$
T	ν	$\sqrt{\frac{\nu}{N}}T_N$, $N=(c+1)\nu+c$	$2\nu/(2-c(\nu+1))$
Double exponential	μ, σ	Double exponential	μ, $\sigma/(c+1)$

for a given estimation functional T and a fixed c, denoted T_c, is obtained as follows.

In theory, at least, we can define a function $\tau_c : \Theta \to \Theta$ by $\tau_c(\theta) = T((G_\theta)_{c,\theta})$, where G_θ is the distribution of the model with parameter θ. This function describes the result of applying the estimator to a model weighted by itself. In many cases, the function τ_c is quite easy to obtain. Often, $\tau_c(\theta)$ is just the parameter of the density proportional to $g^{c+1}(x;\theta)$. Table 1 gives the function τ_c for several common families, where estimation is method of moments.

The robustized estimator T_c is then defined by $T_c(F) = \theta$ where

$$\tau_c(\theta) = T(F_{c,\theta}). \tag{3}$$

The θ that satisfies (3), is obtained by an iterative procedure on the parameters. Namely, let $t^0 = T(F)$ and for $N \geq 0$,

$$t^{N+1} = \tau_c^{-1}(T(F_{c,t^N})), \tag{4}$$

then $T_c(F)$ is a fixed point.

The convergence properties of this procedure are quite reasonable for many useful model families and estimation procedures. In particular, for M-estimators, a value for the asymptotic convergence rate of the iterative procedure can be obtained. For the purposes of this discussion an M-estimator will be defined to be a parameter estimator T, where $T(F)$ is found by solving $E_F[\psi(X, T(F))] = 0$, for a function $\psi : R^d \times R^k \to R^k$.

It follows from (3) that T_c is also an M-estimator obtained by solving for θ in

$$E_F[w(X;\theta)\psi(X,\tau_c(\theta))] = 0. \tag{5}$$

that is, the function for T_c is $\psi_c(x,t) = w(x;t)\psi(x,\tau_c(t))$.

The iterative procedure in (4) finds a fixed point, that is, a solution to $h(t) = t$ for $h(t) = \tau_c^{-1}(T(F_{c,t}))$. The local convergence behavior is characterized by the largest eigenvalue of $h'(t)$ at the solution (Johnson and Riess, 1982, p. 197). If this eigenvalue is less than one, the procedure converges linearly near the fixed point. The smaller the eigenvalue is, the faster the process converges. For M-estimators, the iterating function satisfies $E_{F_{c,t}}[\psi(X,\tau_c(h(t)))] = 0$. Differentiating with respect to t leads to

$$\begin{aligned} h'(t) &= -c\{E_{F_{c,t}}[\psi']\tau_c'(t)\}^{-1}E_{F_{c,t}}[\psi \tfrac{\partial}{\partial t}\log g] &(6)\\ &= c\,B(t)\{I + c\,B(t)\}^{-1}. &(7)\end{aligned}$$

at a fixed point, where

$$B(t) = -E_F[\psi_c']^{-1}E_F[\psi_c \tfrac{\partial}{\partial t}\log g]. \tag{8}$$

Within the model family, i.e. $F = G_\theta$, $E_{G_\theta}[\psi_c(X,\theta)] = 0$, for all θ. Differentiating this equation with respect to θ, leads to $B(\theta) = I$, so that, for F "near" a model, local convergence at a rate of about $r = c/(c+1)$ should be expected.

4. EXPONENT SELECTION

The criterion ρ for making the "best" choice for the exponent c measures the retention of efficiency in using a model G to describe a sample from a distribution F. The purpose of this section is to define ρ and to justify this interpretation.

The matrix B in (8) is related to the influence function. In fact, we have $B(T_c(F)) = E_F[IF(X;T_c,F)s^T(X;T_c(F))]$, where $s(x;t) = \frac{\partial}{\partial t}\log g(x;t)$, the maximum likelihood scores function for the model G_t. If $F = G_t$, then B is the identity. It is reasonable to suppose that the further this matrix is from the identity, then the less alike the two distributions are. This interpretation is even more reasonable because the matrix is related to asymptotic efficiency. The relationship is easier to see if the parameter is one-dimensional.

The asymptotic efficiency of an M-estimator for using a model G to describe a sample from a distribution F is given by $\{E_F(s^2)E_F(IF^2)\}^{-1}$, the reciprocal of the Fisher information times the asymptotic variance. It follows from the Cauchy-Schwartz inequality that B^{-2} is an upper bound for the efficiency. Of course, if F and G are the same, then B^{-2} is one. On the other hand, if B^{-2} were much less than one, then one could say that using G to describe a sample from F is not very efficient. In other words, B^{-2} measures efficiency retention when using G to model F. If the retention is low, then G is not a good model for F. This rationale is the basis for the exponent selection criterion ρ, which is defined to be

$$\rho(c) = B^{-2}(T_c(F)),$$

or its smallest eigenvalue in the multiparameter case. The value of c that maximizes ρ corresponds to the model for which the efficiency retention is the greatest, and that is the model chosen.

It follows readily from (7) that $\rho(c) = (c/r - c)^2$, where r is the largest eigenvalue of h', that is, the convergence rate of the iterative procedure. The convergence rate for a fixed point iteration can be estimated from successive parameter estimates. In particular, if θ^N is the parameter estimate at the N-th and last iteration, the convergence rate can be estimated by,

$$\hat{r} = \frac{|\theta^N - \theta^{N-1}|}{|\theta^{N-1} - \theta^{N-2}|}.$$

In practice, $\rho(c)$ is computed by $\rho(c) = (c/\hat{r} - c)^2$.

It may be somewhat surprising that the convergence rate of a parameter estimation algorithm has a statistical interpretation. A similar circumstance occurs in fitting mixture models, when the EM algorithm is used. The convergence rate of the EM can be interpreted in terms of Fisher information and has been used with some success in assessing the quality of a mixture model fit (Windham and Cutler, 1992).

5. ROBUSTNESS

A quantitative discussion of the robustness of a robustized model fitting procedure is complicated by the fact that the estimate depends on c, the exponent. For example, if the original estimator T is an M-estimator, then for a particular exponent c, T_c is also an M-estimator, and a great deal of mathematical machinery can be used to study robustness. By choosing the estimate that maximizes the criterion ρ, the entire process is no longer an M estimator. At this point the best I am able to do is discuss the robustness of T_c for a fixed c, and, in fact, most of the results apply to M-estimators. The restriction to M-estimators is not that serious, in that the goal is to improve the properties of simple estimators, such as maximum likelihood or method of moments, both of which are M-estimators.

Some characteristics of robustness for a parameter estimator can be investigated with the influence function $\mathrm{IF}(\,\cdot\,; T, F)$ (Hampel et al., 1986). The value $\mathrm{IF}(x; T, F)$ measures the effect on $T(F)$ of adding an observation at x. The supremum in x of the magnitude of the influence function, the gross error sensitivity $\gamma^*(T, F)$, is a measure of the stability of T under small changes in F, hence measures local robustness. Another important measure of robustness is the breakdown point $\varepsilon^*(T, F)$, which describes at what distance from the distribution the estimator no longer gives reliable information. The hope is that an estimator will have a low gross error sensitivity and a high breakdown point. Unfortunately, such basic estimators as maximum likelihood and method of moments often have infinite gross error sensitivity and zero breakdown point, the worst cases.

For an M-estimator T and a fixed c, the influence function for T_c is given by

$$\mathrm{IF}(x; T_c, F) = w(x; \theta)\{\tau_c'(\theta)(I - h'(\theta))\}^{-1}\mathrm{IF}(x; T, F_{c,\theta}),$$

where $\theta = T_c(F)$ and h' is given in (6). The presence of the weight function w makes the gross error sensitivity of T_c is finite for many model families, even when it is not for T itself.

The breakdown point is a global measure of robustness and is perhaps the more important of the two. I have no general results, but the following relation has been useful in specific cases. For a fixed distribution G, if there are constants k and m so that for any distribution F for which

$$|E_G[IF(x; T, F)]|/\gamma^*(T, F) < k, \text{ it follows that } |T(F) - T(G)| < m, \quad (9)$$

then $\varepsilon^*(T, G) \geq k/(1 + k)$. This fact is particularly useful for situations involving scale parameters. For example, it follows readily that for estimating the variance by fitting a univariate normal with known mean using robustized maximum likelihood, the breakdown point satisfies $\varepsilon^* \geq 1/(1 + \max(1, \frac{2(c+1)}{c}e^{-\frac{c}{2(c+1)}-1}))$. In particular, for large enough c the breakdown point is at least one half, but approaches zero as c approaches zero.

Unfortunately, (9) is not always useful, even in some simple cases. For example, for estimating the mean by fitting a univariate normal with known variance using robustized maximum likelihood, (9) does not hold. In fact, the breakdown point may well be zero for this example. For $F = (1 - \varepsilon)G + \varepsilon H$, with G a standard normal, the iterative procedure for the mean has, in general, two fixed points, one near the mean of G and the other near the mean of H. Which of the two is obtained depends in part on where the iterations are started. Starting with the mean of F would be expected to produce the desired result for ε sufficiently small. Using a high breakdown starting point would provide greater assurance, but simulations presented in Section 6 ignore this possibility and the results are still quite good.

6. SIMULATIONS

A feel for the effectiveness of the robustizing procedure can be obtained from simulations. Simulations are particularly important since analytic measures of robustness for the complete robustizing process, including automatically choosing c, are not available.

The simulations fitted a univariate normal to data, where the parameters were the location and scale. These parameters were estimated using using both maximum (MLE) and robustized maximum likelihood (RMF). The maximum likelihood estimates are the sample mean and standard deviation.

The robust estimators mentioned in Section 2 were computed as well.

Two sets of experiments were performed. The first consisted of 500 replications of estimation for a sample of size 100 from a standard univariate normal. The second consisted of 500 replications of estimation for a sample of size 100 from a mixture of two normals, $.95\phi(x; 0, 1) + .05\phi(x; 4, 1)$. The latter data sets contained about five percent contamination, concentrated to the right of the bulk of the data, but not so far to the right that all outliers are obvious. Table 2 shows the mean square errors, standard deviations, and biases for the location and scale, respectively.

These tables suggest that the RMF estimates performed well whether or not contamination was present. When contamination was present, the robustized estimates performed as well as or better than the other robust estimators. On the other hand, the results for uncontaminated data suggest that the robustized estimators may not be quite as efficient as the Huber and bisquare estimators.

Any of these estimates could be used for outlier rejection. Each sample was examined to determine whether a standard outlier rejection test would identify the maximum of the data as an outlier. The test statistic, $(\max\{x_i\} - m)/s$, was used, where m and s, were either the mean and standard deviation or their robustized versions. Values of the statistics were compared to 5% and 1% critical values given in Grubbs and Beck (1972). These critical values are based on the null hypothesis that the data are a sample from a standard normal, m is the sample mean and s is the sample standard deviation. For normal data with no contamination, the maximum

TABLE 2: Simulation results ($\times$ 1000) for normal model fitting.

Location						
	Normal			Contaminated		
	MSE	SD	Bias	MSE	SD	Bias
MLE	10.5	3.2	5.0	59.1	4.0	207.0
RMF	11.3	3.4	5.4	13.5	3.7	11.2
Median	15.4	3.9	2.9	19.8	3.8	71.7
Huber	10.5	3.2	4.6	21.6	3.4	100.4
Bisquare	10.6	3.3	4.4	15.4	3.4	58.4

Scale						
	Normal			Contaminated		
	MSE	SD	Bias	MSE	SD	Bias
MLE	4.8	2.2	-5.4	115.4	4.6	307.1
RMF	6.4	2.5	-9.2	9.1	3.0	3.2
MAD	11.8	3.4	-7.5	17.2	3.7	58.5
Huber-τ	5.8	2.4	-11.8	16.9	3.1	86.5
A-estimate	5.9	2.4	-2.9	10.8	2.9	45.9

likelihood estimates identified the maximum as an outlier 6.4% of the time at the 5% level and 0.6% at 1% level, about as expected. The robustized estimator tended to reject a bit too often, namely, 9.0% at the 5% level and 2.4% at 1% level. For the contaminated data, the MLE performed poorly, rejecting the maximum 85 and 53 percent of the time, suggesting that masking is taking place. Masking was not a problem for the robustized estimates, where the results were 98% at the 5% level and 95% at 1% level. These rates are based on critical values for a situation where the null hypothesis does not strictly apply. An alternative assessment was made using as critical values, the 95-th and 99-th percentiles robustized estimates, obtained from the uncontaminated data. Doing so yields rejection rates for the contaminated data of 97% and 91%.

7. FINAL COMMENTS

The goal of the procedure presented in this paper is to take advantage of the structure of a model family to improve the fit by one of its members. The model is fit to that part of the data that is most compatible with the models in the family.

The method has flexibility, in that it can be easily implemented with a many model families, including multivariate normal, gamma, beta, T, F, and double exponential. The method itself does not implicitly impose assumptions like symmetry in data, the only assumption required is which parametric family to use. Furthermore, the tuning constant can be used to adapt the model family to the data, or unlike in other robust methods, can be chosen optimally.

REFERENCES

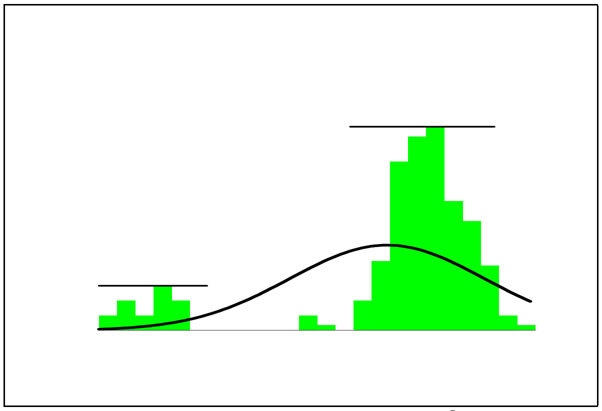
Grubbs, F. E. and Beck, B. (1972) Extension of Sample Sizes and Percentage Points for Significance Tests of Outlying Observations. *Technometrics*, **14**, 847–854.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986) *Robust Statistics: The Approach Based on Influence Functions.* New York: Wiley.

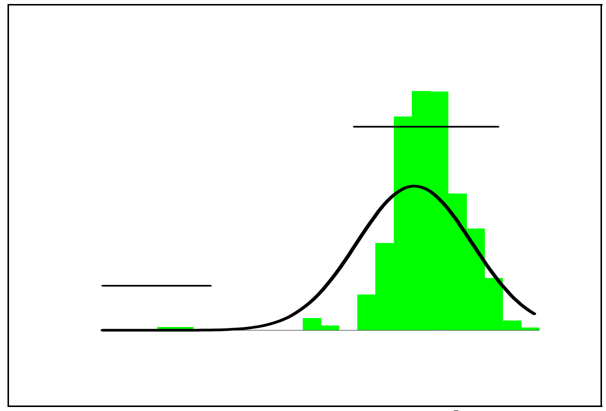
Johnson, L. W. and Riess, R. D. (1982) *Numerical Analysis.* Reading: Addison-Wesley.

Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. (1990) *Numerical Recipes in C.* Cambridge: Cambridge University Press.

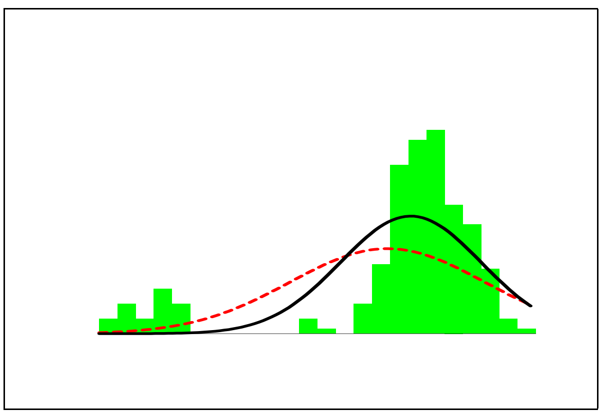
Windham M. P. and Cutler A. (1992) Information Ratios for Validating Mixture Analyses. *Journal of the American Statistical Association*, **87**, 1188–1192.



a. Original data, $\hat{\mu} = -.98$, $\hat{\sigma}^2 = 7.35$

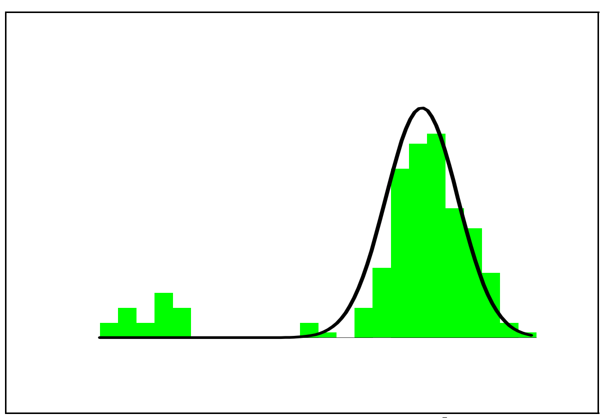


b. Weighted data, $\hat{\mu} = -.33$, $\hat{\sigma}^2 = 2.57$



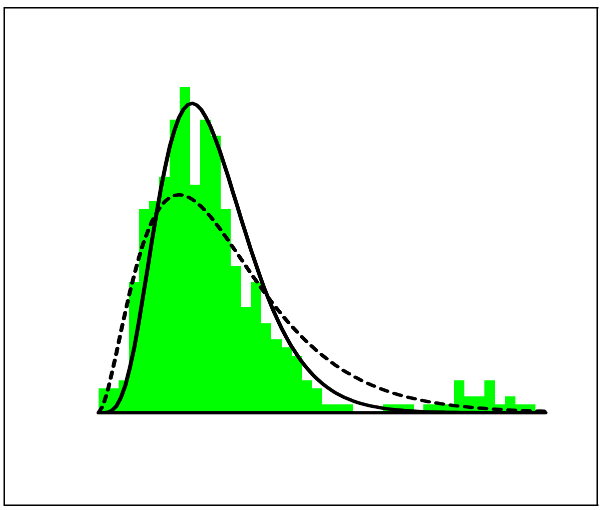
c. Original data, $\hat{\mu} = -.33$, $\hat{\sigma}^2 = 3.85$

Iteration	$\hat{\mu}$	$\hat{\sigma}^2$
0	-.98	7.35
1	-.33	3.85
2	-.11	1.74
3	-.06	1.19
4	-.05	1.06
5	-.04	1.03
6	-.04	1.01



d. Final model, $\hat{\mu} = -.04$, $\hat{\sigma}^2 = 1.01$

FIGURE 1: Robustized maximum likelihood estimation.



c	Shape	Scale	$100\rho(c)$
.0	2.77	4.08	32.7
.2	3.32	3.30	16.0
.4	4.44	2.36	11.7
.6	5.21	1.97	37.4
.8	5.43	1.88	69.8
1.0	5.50	1.86	84.3
1.2	5.53	1.85	86.3
1.4	5.55	1.84	83.5
1.6	5.59	1.83	79.3
1.8	5.62	1.82	74.9
2.0	5.66	1.80	70.8

FIGURE 2: Robustized method of moments for gamma models.