

The Maximum-Error Test

Douglas C. Freesland,¹ Yang Cheng,² and Malcolm D. Shuster³

O hateful error, melancholy's child,
Why dost thou show to the apt thoughts of men
The things that are not?

William Shakespeare (1564–1616)
Julius Caesar, Act V, scene iii

Abstract

The maximum-error test is given a rigorous statistical analysis. It is shown that the test is a poor figure of merit for judging the quality of performance of an attitude estimation algorithm. Two illustrative examples are presented.

Introduction

The maximum-error test (or, better, maximum-absolute-error test) is simply the largest absolute value of the (scalar) error observed in a sequence of simulation trials. This test is found very seldom in the journal literature. In attitude estimation the maximum-error test is encountered more frequently in mission development during the mission software validation and verification process, although it is more common to encounter the three-sigma error. In one of the few journal occurrences in the literature on spacecraft attitude estimation, Markley and Mortari [1] present the maximum error alongside a more transparent figure of merit, the sampled standard deviation of the error. Examples exist in the non-archival literature in which the maximum error is the sole figure of merit presented. The most important characteristic of the maximum observed error is that it is ambiguous as a figure of merit.

¹Chief Technical Officer, ACS Engineering, 7341 Eden Brook Drive, Columbia, Maryland 21046. E-mail: dougfreesland@aol.com.

²Research Assistant Professor, Department of Mechanical and Aerospace Engineering, University at Buffalo, State University of New York, Amherst, New York 14260. E-mail: cheng3@buffalo.edu, website: www.acsu.buffalo.edu/~cheng3.

³Director of Research, Acme Spacecraft Company, 13017 Wisteria Drive, Box 328, Germantown, Maryland 20874. E-mail: mdshuster@comcast.net, website: <http://home.comcast.net/~mdshuster>.

This is because the behavior of this figure of merit turns out to be dominated by its dependence on the number of samples in a simulation test, and because it has a large standard deviation.

The maximum-error test certainly predates the space age and the advent of probability theory. The Gaussian distribution, the basis for modern error analysis, did not come into popular use until the nineteenth century. That distribution, in fact, is not attributable to Gauss but to the Frenchman Abraham de Moivre, a contemporary and close friend of Newton in the eighteenth century (see Appendix A). It is Gauss, however, who is probably most responsible for the promulgation of the normal distribution. It cannot have been long after the publication of Gauss' work on least-squares that the use of the three-sigma confidence interval became popular as containing 99.74 percent of the data (in practice frequently only 99 percent), even if that data were not normally distributed.⁴ The maximum error, the least upper bound of *all* the observed absolute errors, might seem to be an equally descriptive measure of quality. However, it is a very different beast from the three-sigma error bound, even if it seems close, as we shall soon see.

The maximum-error test has never been subjected to a serious quantitative evaluation, at least not within the framework of Astronautics. The present work seeks to remove that lacuna. The presentation below is not specific to attitude parameters, and is almost entirely devoted to simple scalar examples.

The Maximum Observed Error

Consider first an extremely simple example. Let $x^{r.v.}$ be a non-negative (scalar) random variable and let its probability density function (pdf) be denoted by $p_x(x')$. If we are interested in applying the maximum-error test to a variable which can have both non-negative and negative values, then we can consider equivalently the pdf

$$p_x^+(x') \equiv \begin{cases} p_x(x') + p_x(-x') & \text{for } x' > 0 \\ p(0) & \text{for } x' = 0 \end{cases} \quad (1)$$

for non-negative values of x' . Assuming that the domain of x is always restricted to the non-negative axis, the probability function $P_x(x')$ is defined in the usual way

$$P_x(x') \equiv \text{Prob}(\{x < x'\}) = \int_0^{x'} p_x(t) dt \quad (2)$$

If $p_x(x')$ has finite support with least upper bound a , and is uniformly piecewise continuous on $[0, a)$, then

$$P_x(0) = 0 \quad \text{and} \quad P_x(a) = 1 \quad (3ab)$$

Otherwise, we say that $p_x(x')$ has infinite support and, if it is uniformly piecewise continuous on $[0, \infty)$,

$$P_x(0) = 0 \quad \text{and} \quad P_x(\infty) \equiv \lim_{x' \rightarrow \infty} P_x(x') = 1 \quad (4ab)$$

Consider n independent identically-distributed random variables $x_i^{r.v.}$, $i = 1, \dots, n$, with pdf $p_x(x')$, random variables which represent the n simulation test samples, and define⁵

⁴In contrast, working engineers when they speak of the "sigma" of a data set, generally mean the sampled standard deviation. Hence "sigma" and "three-sigma" in general usage in mission support by working engineers not interested in research or rigorous probability theory may not be related by a factor of three.

⁵Naturally, the random variable $x_{\max}^{r.v.}(n)$ is that which has the realization $x'_{\max}(n)$.

$$x_{\max}^{\text{r.v.}}(n) \equiv \max_{1 \leq i \leq n} x_i^{\text{r.v.}} \quad (5)$$

For any probability distribution

$$\text{Prob}(\{x_{\max}(n) < y\}) = [\mathbf{P}_x(y)]^n \quad (6)$$

For a pdf with finite support, since $0 \leq \mathbf{P}(y) < 1$ for $y < a$, it must be that

$$\lim_{n \rightarrow \infty} \text{Prob}(\{x_{\max}(n) < y\}) = 0 \quad \text{for } y < a \quad (7)$$

It follows for a pdf with finite support that

$$\lim_{n \rightarrow \infty} x_{\max}^{\text{r.v.}}(n) = \text{l.u.b. } x \quad \text{w.p. } 1 \quad (8)$$

where “l.u.b.” is an abbreviation for “least upper bound,” and “w.p. 1” for “with probability 1.” Thus, for an infinite number of samples the maximum observed error will converge to the least upper bound of the support of the pdf, no matter what the quality of the results (barring, of course, mistakes in the simulation). In a similar way, for a pdf with infinite support, one can show that

$$\lim_{n \rightarrow \infty} x_{\max}^{\text{r.v.}}(n) = \infty \quad \text{w.p. } 1 \quad (9)$$

As a result, the maximum error when n is large may not be able to say much about the behavior of x .

Illustrative Examples

Uniform Probability Density Function

As an example of a pdf with finite support, consider the uniform distribution on the interval $[0, 1]$, so that $\mathbf{P}_x(x') = x'$. Then

$$\text{Prob}(\{x_{\max}(n) < y\}) = y^n \quad (10)$$

and the pdf of $x_{\max}^{\text{r.v.}}$, $\rho_{x_{\max}(n)}(y)$, is just ny^{n-1} , whence one obtains straightforwardly

$$E\{x_{\max}^{\text{r.v.}}(n)\} = \frac{n}{n+1} \quad \text{and} \quad \text{Var}\{x_{\max}^{\text{r.v.}}(n)\} = \frac{n}{(n+1)^2(n+2)} \quad (11\text{ab})$$

with $E\{\cdot\}$ denoting the expectation and $\text{Var}\{\cdot\}$ the variance. As $n \rightarrow \infty$ we have that $\sigma_{x_{\max}}$, the standard deviation of $x_{\max}^{\text{r.v.}}(n)$, approaches $1/n \rightarrow 0$. Thus, as $n \rightarrow \infty$, $x_{\max}^{\text{r.v.}}(n) \rightarrow 1$ with probability 1.

For this example of a pdf with finite support, we see explicitly that the maximum-error will be a poor criterion for judging the appropriateness of a simulation, because for sufficiently large n the value of the maximum error tends to a universal limit which will be the least upper bound of the support of the pdf.

Gaussian Probability Density Function

As an example of a pdf with infinite support consider a Gaussian random variable with zero mean and unit variance. For a zero-mean Gaussian random variable with standard deviation σ , this is equivalent to studying x/σ . Since we restrict the random variable to non-negative values⁶

$$\rho_x(x') = \sqrt{\frac{2}{\pi}} e^{-(x')^2/2}, \quad 0 \leq x' < \infty \quad (12)$$

⁶To be consistent with our earlier remarks, we should really write not $\rho_x(x')$ but $\rho_x^+(x')$.

and

$$P_x(x') = \operatorname{erf}(x'/\sqrt{2}), \quad 0 \leq x' < \infty \quad (13)$$

with $\operatorname{erf}(z)$ the error function defined as

$$\operatorname{erf}(z) \equiv \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt \quad (14)$$

Thus, $\operatorname{erf}(0) = 0$, and $\operatorname{erf}(\infty) = 1$.

For the Gaussian distribution, it follows that

$$\operatorname{Prob}(\{x_{\max}(n) < y\}) = [\operatorname{erf}(y/\sqrt{2})]^n \quad (15)$$

with the resulting pdf

$$\rho_{x_{\max}(n)}(y) = n[P_x(y)]^{n-1}\rho_x(y) = n\sqrt{\frac{2}{\pi}}e^{-y^2/2}[\operatorname{erf}(y/\sqrt{2})]^{n-1} \quad (16)$$

The expectation of $x_{\max}^{\text{r.v.}}(n)$ in this case cannot be calculated in closed form. However, we can calculate easily the mode of $\rho_{x_{\max}(n)}(y)$, i.e., the value of y at which $\rho_{x_{\max}(n)}(y)$ is a maximum. In the asymptotic limit we anticipate that the expectation of $x_{\max}^{\text{r.v.}}(n)$ will approach this value. From Fig. 1, we see that for n large, $\rho_{x_{\max}(n)}(y)$ is sharply peaked, and, therefore, the expectation of $x_{\max}^{\text{r.v.}}(n)$ will be close to the mode of the pdf. Denoting the mode of $\rho_{x_{\max}(n)}(y)$ by $y_{\text{mode}}(n)$, this quantity satisfies⁷

$$y_{\text{mode}}(n) = (n - 1) \frac{\rho_x(y_{\text{mode}}(n))}{P_x(y_{\text{mode}}(n))} \quad (17)$$

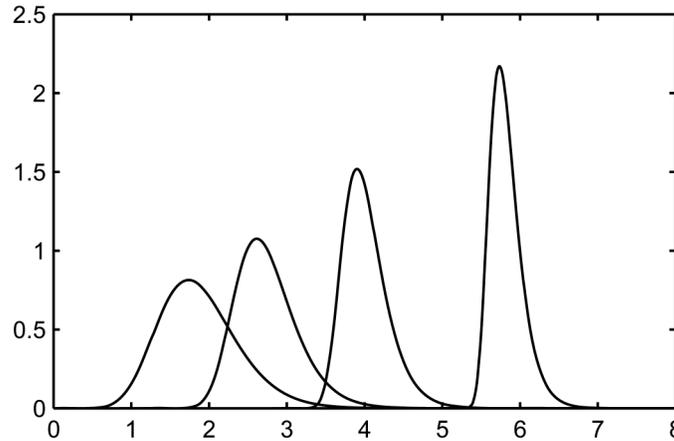


FIG. 1. Probability Density Function for $x_{\max}^{\text{r.v.}}(n)$ for $n = 10, 100, 10,000$ and $100,000,000$. The location of the peak increases with increasing n . The single-sample random variable is assumed to be Gaussian with mean zero and unit variance. The abscissa of this graph is the non-random variable y , and the ordinate is $\rho_{x_{\max}(n)}(y)$.

⁷We write $y_{\text{mode}}(n)$ and not $x_{\text{mode}}(n)$, because there is no random variable $x_{\text{mode}}^{\text{r.v.}}(n)$. The quantity $y_{\text{mode}}(n)$ is simply a characteristic of the function $\rho_{x_{\max}(n)}(y)$; it is not related to the realization of any random variable. We could have written this quantity as $x_{\text{mode}}(n)$, but we chose not to do so.

hence,

$$1 + \sqrt{\frac{\pi}{2}} y_{\text{mode}}(n) e^{(y_{\text{mode}}^2(n)/2)} \text{erf}(y_{\text{mode}}(n)/\sqrt{2}) = n \quad (18)$$

For n large, equation (18) can be solved crudely as

$$y_{\text{mode}}(n) \approx \sqrt{2 \ln n} \quad (19)$$

showing that $y_{\text{mode}}(n) \rightarrow \infty$, and $\Delta y_{\text{mode}}(n)/\Delta n \rightarrow 0$, as $n \rightarrow \infty$.

We can write the variance of $x_{\text{max}}^{\text{r.v.}}(n)$ in the asymptotic limit as⁸ [2]

$$\begin{aligned} \text{Var}\{x_{\text{max}}^{\text{r.v.}}(n)\} &= - \left[\frac{d^2}{dy^2} \ln[\rho_{x_{\text{max}}(n)}(y)] \right]_{y=y_{\text{mode}}}^{-1} \\ &= \left[1 + \frac{n}{n-1} y_{\text{mode}}^2(n) \right]^{-1} \quad \text{for } n > 1 \end{aligned} \quad (20)$$

Thus, for $n \gtrsim 20$ the standard deviation of $x_{\text{max}}^{\text{r.v.}}(n)$ is approximately

$$\sigma_{x_{\text{max}}(n)} \approx \frac{1}{\sqrt{1 + y_{\text{mode}}^2(n)}} \quad (21)$$

Clearly, $\text{Var}\{x_{\text{max}}^{\text{r.v.}}(n)\} \rightarrow 0$ as $n \rightarrow \infty$, and $x_{\text{max}}^{\text{r.v.}}(n) \rightarrow \infty$ with probability 1 in that limit.

Thus, the value of the maximum error is really not significant for judging a simulation for the cases of a pdf with infinite support, since the maximum error can attain any value provided that a sufficient number of samples are included in the test.⁹ In addition, the standard deviation of the maximum error is quite large compared to other statistical measures (e.g., the mean and the variance) and decreases extraordinarily slowly with increasing n . Furthermore, we examined above only the example of a single random Gaussian random variable. In general, attitude systems contain many (vectorial) error sources all characterized by different variances and covariances. The simple formula of equation (18) will certainly not apply in general, and the specific nature of the increase of the expectation value of $x_{\text{max}}^{\text{r.v.}}(n)$ will be more complex and, perhaps, not even Gaussian. What will not change is the fact that by taking a sufficient number of samples, $x'_{\text{max}}(n)$ can be made to exceed any value.

In Table 1, we show the values of n corresponding to different values of $y_{\text{mode}}(n)$ rather than the converse, which would have been much more difficult to calculate. We note from the table and from equation (19) that $y_{\text{mode}}(n)$, although it must become infinite in the limit that $n \rightarrow \infty$, grows very slowly with increasing n . For n very large, the value of $y_{\text{mode}}(n)$ is not a very useful figure of merit. In the table the number $a \times 10^b$ is denoted by ‘‘a E b.’’ All entries have been rounded to two significant figures.

Figure 2 shows a histogram of 100 samples of $x_{\text{max}}^{\text{r.v.}}(n)$ for $n = 1,700,000$, chosen so that $y_{\text{mode}}(n)$ would be close to 5.0. Our sample showed a sampled mean of the hundred values of $x'_{\text{max}}(n)$ of 5.09 and a sampled standard deviation of 0.19 compared with a theoretical mean of 5.0 and a theoretical standard deviation of 0.19.

⁸It cannot have escaped the reader’s notice that the techniques applied here are very similar to those of maximum-likelihood estimation.

⁹Extremely large or small values of $x_{\text{max}}(n)$, of course, will indicate catastrophic problems with the analysis, software, or mission geometry associated with the simulation, even if the ‘‘resolution’’ of $x_{\text{max}}(n)$ is not fine enough for detailed evaluation. If one has 100 samples of a test variable $x \sim \mathcal{N}(0, 1)$, and one finds values of $x_{\text{max}}(n)$ of 1000. or 0.001, one knows that something is very wrong.

TABLE 1. Number of Samples versus Modal Maximum Error for a Zero-Mean Gaussian Random Variable with Unit Variance

$y_{\text{mode}}(n)$	n
1.	3
2.	20
3.	340
4.	15,000
5.	1,700,000
7.	3.8 E+11
10.	6.5 E+22
15.	1.4 E+50
20.	1.8 E+88
25.	1.6 E+137
30.	1.0 E+197
35.	4.4 E+267
40.	1.4 E+349
45.	3.0 E+441
50.	4.6 E+544

This should be compared with the mean of the sampled standard deviation of the example (as the square root of the sampled variance) of 1.000. The standard deviation of the sampled standard deviation of the x'_k , $k = 1, \dots, 1,700,000$ was 0.00055, showing that the standard deviation of the 1.7 million x' is a much more sensitive statistic than the maximum error, which was to be expected, and which is surely also demonstrated by Table 1. The highest sampled value of $x'_{\text{max}}(n)$ corresponds to a 3.6-sigma event (for $x'_{\text{max}}(n)$). Two sampled values were beyond the three-sigma boundary. These correspond to more than 5.6-sigma events for x , at which point the quality of the simulation of the normal distribution begins to become suspect. Frequently, a normal distribution is simulated by averaging only 12 samples of a uniform distribution [3].

That only two values of $x'_{\text{max}}(n)$ in our experiment lie beyond the three-sigma boundary was not unexpected. A three-sigma event has a probability of approximately

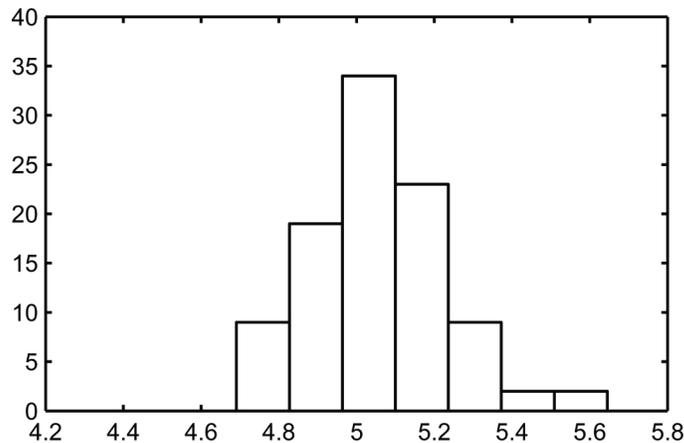


FIG. 2. Distribution of maximum error for 100 tests with 1.7 million samples in each test. The abscissa is the value of the sample $x'_{\text{max}}(n)$, and the ordinate is the number of samples.

0.25 percent. Thus, with 100 samples from a Gaussian distribution, we would expect on the average that 1/4 of a sample will be three-sigma events. That we obtain slightly more than that may be attributed to the non-Gaussian nature of the distribution of $x_{\max}^{r.v.}(n)$, illustrated also by the lack of symmetry about the mean, or to a chance occurrence. For 1.7 million tests, we would expect the largest sampled value of $x_{\max}^{r.v.}(n)$ to be a 4.7-sigma event. (Note that the value for sigma depends on n , but the maximum maximum-error sample as a multiple of sigma observed in N simulations of n samples of x depends more on the number of samples of $x_{\max}^{r.v.}(n)$ than on the underlying random variable $x^{r.v.}$.)

Alternative Figures of Merit

A much more informative test is the comparison of the sampled mean and covariance matrix of the simulation compared with its predicted¹⁰ value, assuming this can be known separately from the simulation. Such figures of merit are not totally dependent on the value of outliers.

In the case where the subject of the study is three-axis attitude estimation, the preferred attitude error parameterization is $\Delta\xi^*$, the Cartesian attitude error increment defined [4] as $\Delta\xi$ where

$$A^* = \delta A(\Delta\xi^*) A^{\text{true}} \quad (22)$$

with A the direction-cosine matrix, A^* its estimator, and $\delta A(\Delta\xi)$ the direction-cosine matrix parameterized by the rotation vector $\Delta\xi$ [3], in this case infinitesimal. As usual, the asterisk indicates the estimator, a random variable whose realization is the estimate.

For spin-axis attitude estimation [5, 6] the obvious choice is $\Delta\hat{\mathbf{n}}^*$, the error in the spin-axis vector

$$\Delta\hat{\mathbf{n}}^* = \hat{\mathbf{n}}^* - \hat{\mathbf{n}}^{\text{true}} \quad (23)$$

The covariance matrix $P_{\xi\xi}$ has the good quality of being covariant under a change of basis of the spacecraft body coordinates. The covariance matrix $P_{\hat{\mathbf{n}}\hat{\mathbf{n}}}$ has this same quality under a change of basis of the space coordinate system, but the disadvantage of being singular, since the spin-axis attitude can only be of rank 2. Reference [6] has studied two possibilities for a two-dimensional error, one of which, the incremental vector, is similar to $\Delta\xi$ in concept. This latter error, $\tilde{\mathbf{e}}$, relies on a basis $\{\hat{\mathbf{a}}^{\text{true}}, \hat{\mathbf{b}}^{\text{true}}, \hat{\mathbf{n}}^{\text{true}}\}$, where $\hat{\mathbf{n}}^{\text{true}}$ is the true spin-axis vector and the remaining two vectors, which span the plane of the spin-axis estimation errors, may be chosen to be fixed in the spacecraft body, since the spin-axis is always assumed to be fixed in the spacecraft body. (If it is not, then there is little reason to estimate the spin-axis attitude.) The 2×2 *body-referenced* covariance matrix is then, barring problems of observability, full rank.

In orbit determination, the spacecraft position and velocity, being Cartesian vectors, require no special attention in the definition of errors.

If a single scalar figure of merit is desired, then the obvious choice is the root-mean-square error.¹¹ If a practical model for the covariance matrix of the estimate

¹⁰The word "predicted" here does not mean predicted in the sense of the Kalman filter but only predicted by the mathematics. This work makes no specific reference to the Kalman filter.

¹¹The mean absolute error is the sampled value corresponding to $E\{|\Delta x|\}$ and has the disadvantage of not being easily relatable to the model covariance matrix.

error is available, then $\Delta \mathbf{x}^T P_{xx}^{-1} \Delta \mathbf{x}$ for P_{xx}^{-1} full rank, provides a figure of merit whose statistics are already known. If $\Delta \mathbf{x}$ has dimension n , then this variable should have a χ^2 distribution with n degrees of freedom.

Two Examples from Attitude Estimation

A Semi-Quantitative Example

Let us consider an illustrative thought experiment. Suppose that we are using the maximum-error test to compare the QUEST and TRIAD algorithms [7] to determine which is better. Let the single-sample error be defined as the largest of the absolute values of the three components of the attitude error vector. The TRIAD algorithm truncates the data, effectively discarding one component of the second observation vector. The effect of this deletion on the statistics of the TRIAD algorithm has been studied in detail [8]. Suppose, that in the entire test data set, say, 1.7 million frames of data, each consisting of two observed directions, there was one bad frame with one bad vector with one bad component, ten-sigma bad, which just happened to have been in the component discarded by the TRIAD algorithm. Then, on the basis of the maximum-error test of, say, the magnitude of the attitude error, *for 1.7 million (!) frames of simulated data*, one would conclude that the TRIAD algorithm performed well and that the QUEST algorithm was simply horrible. This is obviously a worst case. A much less improbable event (or events) of this nature can still make the TRIAD algorithm seem superior to the QUEST algorithm. Our example might seem far-fetched, but it is not so. If we really created 1.7 million frames of simulated data, then from Table 1, we anticipate that we will have many data which are around $5\text{-}\sigma$ events. If it is only one component of the attitude error vector that is bad, then the probability that the worst of six components will be the one discarded by the TRIAD algorithm is $1/6$. This is not a microscopically small probability. It is the reliance on only the six components of the worst measurement pair which is the Achilles' heel of the maximum-error test as an analysis tool.

On the other hand, we would never make such a claim based on a comparison of the sampled attitude-error covariance matrices for the two algorithms for the same test data, in which the bad frame would have been weighted out by 16,999,999 other frames and would change the sampled QUEST covariance matrix by quantities of order $10^{-7}\sigma^2$, a truly negligible amount. Q.E.D.

A More Quantitative Example

Even without such treacherous outliers the maximum-error is still a figure of merit of limited usefulness. In a second thought experiment, suppose again that we are trying to make a relative evaluation of the TRIAD and QUEST algorithms by examining ε_k , the square of the attitude estimation error for each frame k of data for the two algorithms, and the data consists of 1.7 million frames of data with each frame consisting of two perpendicular unit-vector measurements corrupted by noise of mean zero and variance σ^2 . Then the ε_k ,

$$\varepsilon_k \equiv |\xi_k^*|^2, \quad k = 1, \dots, 1,700,000 \quad (24)$$

each has mean and variance¹²

¹²Note that $\varepsilon_k^{\text{TRIAD}} \sim \sigma^2 \chi^2(3)$, but the same is not true for $\varepsilon_k^{\text{QUEST}}$, because the three (independent) components of ξ_k^{QUEST} in our example do not all have the same variance [7]. Hence, the factor in equation (25d) is $\sqrt{9/2}$ rather than $\sqrt{5}$.

$$\mu_{\varepsilon}^{\text{TRIAD}} = 3 \sigma^2, \quad \sigma_{\varepsilon}^{\text{TRIAD}} = \sqrt{6} \sigma^2 = 2.45 \sigma^2 \quad (25ab)$$

$$\mu_{\varepsilon}^{\text{QUEST}} = (5/2) \sigma^2, \quad \sigma_{\varepsilon}^{\text{QUEST}} = \sqrt{9/2} \sigma^2 = 2.12 \sigma^2 \quad (25cd)$$

We have chosen the square of the absolute value of the attitude-estimation error rather than the absolute value because of the simplicity of equations (25).

For the TRIAD algorithm, noting equation (21) and Table 1, the maximum error $\varepsilon_{\max}^{\text{TRIAD}}$ will have an error level of roughly $\sigma_{\varepsilon}^{\text{TRIAD}}/5 = (\sqrt{6}/5) \sigma^2 = 0.49 \sigma^2$, and that for the maximum error $\varepsilon_{\max}^{\text{QUEST}}$ for the QUEST algorithm will be roughly $\sigma_{\varepsilon}^{\text{QUEST}}/5 = (\sqrt{9/2}/5) \sigma^2 = 0.42 \sigma^2$. The discriminator for the relative evaluation of the QUEST and TRIAD algorithms will be $\varepsilon_{\max}^{\text{TRIAD}} - \varepsilon_{\max}^{\text{QUEST}}$, which will have an error level of

$$\sigma_{\text{max-err-discriminator}} = \begin{cases} \sqrt{21/2} \sigma^2/5 = 0.65 \sigma^2, & \text{for the } \varepsilon^{\max} \text{ uncorrelated} \\ \sqrt{3/2} \sigma^2/5 = 0.24 \sigma^2, & \text{for the } \varepsilon^{\max} \text{ correlated} \end{cases} \quad (26)$$

The presence or lack of correlation will depend on whether $\varepsilon_k^{\text{TRIAD}}$ and $\varepsilon_k^{\text{QUEST}}$ attain their maximum value for the same frame or different frames of data. With such a discriminator we must be able to detect the positivity of

$$\sigma_{\varepsilon}^{\text{TRIAD}} - \sigma_{\varepsilon}^{\text{QUEST}} = 0.33 \sigma^2 \quad (27)$$

The discriminator has an error level which is comparable to the quantity it is supposed to detect. To this we must compare the discriminator based on the difference of the sampled means of the errors $\varepsilon_k^{\text{TRIAD}}$ and $\varepsilon_k^{\text{QUEST}}$, which has an error level on the order of

$$\sigma_{\text{sampled-mean}} \approx \sigma^2 \sqrt{2/1700000} \approx 0.001 \sigma^2 \quad (28)$$

Clearly, if one wishes to detect the sign of a quantity of magnitude $0.33 \sigma^2$ a measurement with an error level on the order of $0.001 \sigma^2$ is superior to a measurement with an error level of $0.24 \sigma^2$ or $0.65 \sigma^2$. Q.E.D.

Discussion and Conclusions

The maximum error has been shown to be a poor indicator of the quality of performance of an estimator, since its value has as much to do with the number of samples in the simulation as with the performance of the estimator. Perhaps, the most unfortunate characteristic of the maximum-error test is that its result becomes less indicative of algorithm performance as the amount of data *increases*. Clearly, for very large samples, the maximum error will be a very unusual and unrepresentative value of the error. If one has 10,000 samples, does it makes sense to examine only the single least representative sample and pay no attention to the other 9,999? A reasonable test would examine the majority of results. Mission designers with good reason specify an attitude system by giving the three-sigma value of the allowable error, by which they usually mean the largest error attained by 99 percent of the cases, and not by the elusive maximum error. They do not specify mission requirements in terms of the most unusual values of the error.

Although the maximum error is both less sensitive and more noisy than other statistics (for example, the sampled mean and the sampled variance), nonetheless, engineers taking a first look at a print-out of the results of a simulation test are most likely (and with good reason) to look first for the largest simulated error, because it is easier to find this number than to compute the sampled mean and covariance of the results. The maximum-error test is thus not without value. Our message is not

that one must always shun the maximum-error test, but rather that its value is at best semi-quantitative. It is useful, because it is quick. It is a first-look tool. However, it is also dirty, and for serious analysis it is not sufficient.

In order that the maximum error be meaningful *quantitatively*, a predicted maximum error is needed to which the sampled maximum error can be compared. In general, for a real system there are many error sources and the system is multi-dimensional and nonlinear. Since maximum error is dependent only on the tails of the probability density functions, a Gaussian approximation may not be adequate even if the input errors are Gaussian, because the error sources will not all have the same variance.¹³ Furthermore, the simulation of a Gaussian random variable by the computer may not be adequate for 5-sigma events.¹⁴ From a practical point of view, the only way to predict the maximum error is by simulation of the errors themselves, which makes no sense at all, because the need for a predicted value for the maximum error is to test this very simulation. The idea is even more ridiculous, because the likely non-Gaussian nature of the tails of the probability density functions generally are not even known. Thus, at best one can compare the sampled maximum error only with vague intuition, which may be unreliable. This inadequacy is in addition to that of the high error level.

As an example of a quantitative comparison by proponents of the maximum-error test, we should examine Table 3 of reference [1]. The next to the last column gives the sampled and maximum boresight error (estimate to truth). It is obvious that the variation of the maximum error is much greater than the root-sum-square error.

For the mean and covariance matrix of the estimation errors, on the other hand, one often has a ready formula, as one does for the TRIAD and QUEST algorithms [7], or at worst one can compute the Fisher information matrix numerically. In both cases one is in the region where the Gaussian approximation is reasonable, that is, in the central portion of the bell curve. Thus, there are theoretical values readily at hand that can be compared with a sampled mean and covariance matrix. For the maximum error, except for the trivial examples in this article, one can do none of this. As an example of reasonable error analysis one can cite reference [6], which compares several spin-axis attitude estimation algorithms not with 1.7 million frames of data but with one hundred. The prosecution rests.

Sic transit testificatio errore maximo.

By the same token, while the three-sigma bound, defined, say, as the least upper bound of 99.73 percent of the absolute-error data, or an equivalent one-sigma bound defined similarly as the least upper bound of 68.27 percent of the data, are useful quick figures of merit but are far from the best. For the three-sigma bound, the accuracy will depend on how many points lie outside the three-sigma bound. This includes only 0.26 percent of the data. Thus, the error level of this quantity must be 20 times larger than that for the sampled variance. This is better than the maximum-error test, provided that there are very many times 400 error samples, but still not very good. The one-sigma-bound test would make use of 31.73 percent of the

¹³A Gaussian asymptotic limit is possible under the less stringent requirements of the Lindeberg or Lyapunov central limit theorem [9], although convergence to a Gaussian limit may not be as rapid.

¹⁴Typically, a Gaussian random variable of mean zero and variance σ^2 is simulated as 12 samples of a uniform random variable on the interval $[-\sigma/2, \sigma/2]$ [3]. Thus, the pdf for arguments which differ from the mean by more than 6σ is modeled as zero, and $p_x(x'|\mu, \sigma^2)$ for $|x' - \mu| > 4\sigma$ may be greatly mismodeled for our purposes.

data, which would make the error level 1.8 times larger than that for the sampled variance.¹⁵ The present writers always quick-check a simulation by verifying that roughly one-third of the estimate errors lie outside the expected one-sigma confidence bounds, but not as an analysis tool.

Acknowledgment

The authors are grateful to F. Landis Markley of NASA and to two anonymous reviewers for helpful comments and criticisms.

References

- [1] MARKLEY, F. L. and MORTARI, M. "Quaternion Attitude Estimation Using Vector Measurements," *The Journal of the Astronautical Sciences*, Vol. 48, Nos. 2 and 3, April–September 2000, pp. 359–380.
- [2] SORENSEN, H. W. *Parameter Estimation*, Marcel Dekker, New York, 1980.
- [3] CODY, W. J. Jr. and WAITE, W. *Software Manual for the Elementary Functions*, Prentice-Hall, Englewood Cliffs, New Jersey, 1980.
- [4] SHUSTER, M. D. "A Survey of the Attitude Representations," *The Journal of the Astronautical Sciences*, Vol. 41, No. 4, October–December 1993, pp. 439–517.
- [5] SHUSTER, M. D. "Efficient Algorithms for Spin-Axis Attitude Estimation," *The Journal of the Astronautical Sciences*, Vol. 31, No. 2, April–June 1983, pp. 237–249; [5e] Errata, Vol. 51, No. 1, January–March 2003, p. 121.
- [6] TANYGIN, S. and SHUSTER, M. D. "Spin-Axis Attitude Estimation," *The Journal of the Astronautical Sciences*, Vol. 55, No. 1, January–March 2007, pp. 107–139.
- [7] SHUSTER, M. D. and OH, S. D. "Three-Axis Attitude Determination from Vector Observations," *Journal of Guidance and Control*, Vol. 4, No. 1, January–February 1981, pp. 70–77.
- [8] SHUSTER, M. D. "The TRIAD Algorithm as Maximum-Likelihood Estimation," *The Journal of the Astronautical Sciences*, Vol. 54, No. 1, January–March 2006, pp. 113–123.
- [9] FELLER, W. *Introduction to Probability Theory and its Applications*, Vol. II, second edition, John Wiley and Sons, New York, 1966, pp. 256–259.
- [10] DE MOIVRE, A. *Approximatio ad Summum Terminorum Binomii $(a + b)^n$ Seriem Expansi*, Proceedings of the Royal Society, November 12, 1733.
- [11] DE MOIVRE, A. *The Doctrine of Chances: A Method of Calculating the Probabilities of Events in Play*, A. Millar in the Strand, London, 1718, 1738, 1756.
- [12] LAPLACE, P.-S. *Théorie analytique des probabilités*, Volume VII of *Œuvres complètes de Laplace*, 14 volumes, l'Académie des sciences, Gauthier-Villars, Paris, 1878–1912.
- [13] LAPLACE, P.-S. *Mécanique Céleste*, 5 volumes, Paris 1799–1825.
- [14] ADRAIN, R. "Research Concerning the Probabilities of the Errors which Happen in Making Observations, &c." *The Analyst, or Mathematical Museum*. Vol. I, Article XIV, pp. 93–109. Philadelphia: William P. Farrand and Co., 1808.
- [15] GAUSS, K. F. J. *Theoria Motus Corporum Coelestium in sectionibus conicis solem ambientium*, Verlag F. Perthes and I. H. Besser, Hamburg, 1809 (esp. pp. 205–224); English translation: DAVIS, C. H. *Theory of the Motion of the Heavenly Bodies Moving About the Sun in Conic Sections*, Little Brown and Company, 1857, reprint: Dover Publications, Mineola, New York, 1963.

Appendix A: Historical Notes

The normal distribution was first introduced by Abraham de Moivre in an article in 1733 [10], which was reprinted in 1738 in the second edition of his *The Doctrine of Chances* [11] in the context of approximating certain binomial distributions for large powers. His result was extended by Laplace in 1812 in his book *Analytical Theory of Probabilities* [12], and is now called the theorem of de Moivre-Laplace.

¹⁵The numbers 20 and 1.8 are the square roots of the relative number of samples participating in the test.

De Moivre, whose book quickly became very popular among gamblers, and is best known for de Moivre's formula

$$(\cos \theta + i \sin \theta)^n = \cos n\theta + i \sin n\theta$$

is also the discoverer of the closed algebraic formula for the Fibonacci Numbers. The approximation for the factorials, wrongly attributed to Stirling, is also to be found in De Moivre's 1733 article.

De Moivre's work was published mostly in Latin or English, the consequence of his flight from France following the revocation in October 1685 by Louis XIV of the Edict of Nantes, which had established religious freedom for Protestants in France almost a century earlier. Although the revocation did not provoke the atrocities of the religious wars of the previous century, which had led Henry IV to proclaim the edict in 1598, it led to the demolition of protestant churches and the seizure of protestant private property in a manner not totally unlike that of the German *Kristallnacht* of 1938. The revocation resulted in the flight of French Protestants to Great Britain, the Dutch Republic, Germany, and Switzerland, creating a brain drain similar to that of continental Europe under the Nazi regime. Among the fall-out was the demise of the French watch industry, Huguenot for the most part, which largely moved from Besançon in Burgundy to Geneva, Switzerland. Among the Huguenots who eventually found their way to America was Apollon Rivoire, father of the silversmith and patriot Paul Revere.

It is Laplace whom we must thank also for the first great treatise applying calculus to the study of Celestial Mechanics [13]. Newton may have used the Calculus to derive for himself the results in his *Principia Mathematica*, but that work contains only geometrical proofs and development.

The normal or "Gaussian" distribution is often wrongly attributed to Adrain and Gauss. The American Robert Adrain first employed the normal distribution in his work on least squares [14]. Carl Friedrich Gauss, in 1795 at the age of 18, had discovered the fundamentals of least-square analysis, but did not publish this work. His work with the normal distribution in least-square estimation did not appear until the publication of Volume II of his *Theory of Motion* [15] in connection with his solution of the Ceres problem in 1801.