

Automatic Propagation of Uncertainties

Bruce Christianson¹ and Maurice Cox²

¹ University of Hertfordshire
Hatfield, England, Europe
B.Christianson@herts.ac.uk

² National Physical Laboratories
Teddington, England, Europe
Maurice.Cox@npl.co.uk

Summary. Motivated by problems in metrology, we consider a numerical evaluation program $y = f(x)$ as a model for a measurement process. We use a probability density function to represent the uncertainties in the inputs x and examine some of the consequences of using Automatic Differentiation to propagate these uncertainties to the outputs y . We show how to use a combination of Taylor series propagation and interval partitioning to obtain confidence intervals and ellipsoids based on unbiased estimators for means and covariances of the outputs, even in the case where f is sharply non-linear, and even when the level of confidence required makes the use of Monte Carlo techniques computationally problematic.

Key words: Confidence interval, covariance, ellipsoid, GUM, kurtosis, metrology, singularities, Taylor series, unbiased estimators, validation.

1.1 Introduction

Often we have a program which calculates the numerical values $y = f(x)$ of some outputs y from the values of the inputs x . For many applications the values of the inputs are not known with certainty. This may be because of a deficit in our knowledge, corresponding perhaps to indeterminacies in the measurement process, or it may be because the input values are themselves representatives of a population with a non-zero variance. These possibilities correspond naturally to the Bayesian and frequentist viewpoints respectively³.

The evaluations of such uncertainties in variable values are referred to respectively as Type B and Type A evaluations of uncertainty in Clauses 2.3.3 and 3.3.5 of the enormously influential methodology for uncertainty evaluation

³ As tool-writers we hope to design our product in such a way as to satisfy both camps that we have performed the correct calculation, and thus to escape involvement in the discussions of the interpretation and significance of the result.

set out in the “Guide to the Expression of Uncertainty in Measurement” [8], published by ISO and popularly known as the GUM.

We often wish to model the uncertainty in the values of the independent variables (inputs x) and to obtain corresponding estimates of the uncertainties in the values of the dependent variables (outputs y).

Interval analysis is one technique which may be used to do this. However in many applications, certainty is either not to be had, or comes at too high a price. For example, the value of a variable drawn from a normal distribution $N(\mu, \sigma)$ is theoretically unbounded, but in practice the confidence interval

$$[\mu - 5\sigma, \mu + 5\sigma]$$

will almost certainly suffice⁴. In other cases, certainty is not desired: for many applications in metrology, mean-centred 95% confidence intervals for the outputs are the information of primary interest.

Direct application of interval analysis to confidence intervals is not straightforward. Input value uncertainties are often correlated as a result of the processes used to obtain them, and we frequently desire to exploit correlation information about output uncertainties. In this paper we use a probability density function (pdf), as in the GUM, to model the input uncertainties, and we examine some of the consequences of using Automatic Differentiation (AD) to propagate these uncertainties to the outputs.

The rest of this paper is organized as follows: in the next section we look at the case where f is linear, corresponding to Clause 8 of the GUM, and introduce the use of multivariate truncated Taylor series. In Sect. 3 we compare and contrast this approach with conventional interval analysis. In Sect. 4 we extend our Taylor series approach to cope with moderate non-linearities in f and show how to obtain unbiased estimators for uncertainty parameters in this case. Implementation issues are considered in Sect. 5. This section also considers how to truncate and partition distributions in order to cope with poles and other artifacts of severe non-linearity. In Sect. 6 we show how AD can be used to validate hypotheses about output distributions, and to construct confidence intervals at various levels of probability under a hierarchy of such assumptions. The final section includes some prospects for the future.

1.2 Linear Models

Many approaches to uncertainty modelling assume that the evaluated function f (known in the GUM as the *model* of a measurement) is linear, or very

⁴ There always comes a point where events within the model become sufficiently unlikely, relative to significant events deliberately ignored by the model, that they can also prudently be disregarded. The majority of metrological analyses, for example, do not consider the effect of the measuring apparatus being struck by a very small meteor at the crucial moment.

nearly linear, at least over sufficiently long intervals surrounding the anticipated values for the inputs⁵.

Under this assumption, the straightforward approach is to represent all the program variables v_i as multivariate first order Taylor series, so that the variable

$$v = (v^{(0)}, v^{(1)}, \dots, v^{(n)}) \quad \text{represents} \quad v^{(0)} + v^{(1)}\zeta_1 + \dots + v^{(n)}\zeta_n ,$$

where the ζ_i are independent random variables with zero mean and unit variance, so that

$$E(\zeta_i) = 0; \quad E(\zeta_i\zeta_j) = \delta_{ij} ,$$

where E denotes expectation.

Usually we will take ζ_i from either the normal distribution $N(0, 1)$ or from the student-t distribution with the appropriate number of degrees of freedom for the number of measurements involved. Other distributions including the uniform and the logarithmic distribution are also possible and are considered in what follows.

1.2.1 Uncorrelated Inputs

In the simplest case, where the uncertainties in the input values to the calculation are uncorrelated, we initialize the input variables x_i by setting

$$x_i = x_i^{(0)} + x_i^{(i)}\zeta_i ,$$

where

$$x_i^{(0)} = E[x_i] \quad \text{and} \quad \left(x_i^{(i)}\right)^2 = V[x_i] = E[(x_i - E[x_i])^2]$$

are the mean and variance respectively of x_i .

1.2.2 Correlated Inputs

In the more general case where the input uncertainties are correlated, we set

$$x_i = x_i^{(0)} + \sum_j x_i^{(j)}\zeta_j ,$$

where $x_i^{(j)} = r_{ij}$ are chosen so that $R = [r_{ij}]$ is the lower triangular decomposition of the input covariance matrix

$$V = RR^T = Cov[x_1, \dots, x_n].$$

⁵ The GUM makes this assumption explicitly [8, Clause G.6.1] as a basis for the uncertainty budgeting (evaluation and expression) procedure set out in Clause 8 of the GUM.

In this case we have $E[x_i] = x_i^{(0)}$ and

$$\begin{aligned} \text{Covariance}(x_i, x_j) &= E[(x_i - E[x_i])(x_j - E[x_j])] \\ &= E\left[\sum_{k,\ell} x_i^{(k)} x_j^{(\ell)} \zeta_k \zeta_\ell\right] = \sum_{k,\ell} x_i^{(k)} x_j^{(\ell)} E[\zeta_k \zeta_\ell] \\ &= \sum_k x_i^{(k)} x_j^{(k)} = \sum_k r_{ik} r_{jk} = v_{ij} \quad , \end{aligned}$$

since by independence $E[\zeta_k \zeta_\ell] = \delta_{k\ell}$.

1.2.3 Single Output

Automatic Differentiation [2, 6, 11] can be used to propagate the numerical coefficients of the Taylor terms through the calculation $y = f(x)$. If the mainstream GUM assumptions [13] of a linear function model are satisfied, then to the required degree of approximation we will have

$$y = y^{(0)} + \sum_j y^{(j)} \zeta_j \quad ,$$

with the mean and variance of y being given by

$$E[y] = y^{(0)} \quad \text{and} \quad V[y] = \sum_j \left(y^{(j)}\right)^2 \quad ,$$

respectively. In the case where the ζ_i are normal, the pdf for y is also normal, and the calculated value for $V[y]$ can be used directly to construct the required confidence interval centred upon $y^{(0)}$.

1.2.4 Multiple Outputs

For a linear model with several outputs AD gives

$$y_i = y_i^{(0)} + \sum_j y_i^{(j)} \zeta_j \quad ,$$

so

$$E[y_i] = y_i^{(0)} \quad , \quad \text{Cov}(y_i, y_j) = \sum_k y_i^{(k)} y_j^{(k)} \quad .$$

These values can be used to construct the required confidence ellipsoid; the values $y_i^{(k)}$ effectively give the covariance of the outputs in a factored form which in the normal case can be used to construct the ellipsoid directly.

1.3 Contrast with Interval Analysis

This section describes a simple linear example, intended to illuminate some differences between the approach taken in this paper and that of conventional interval analysis.

Consider the case where the uncertainties in the inputs x_1 and x_2 are modelled by independent (uncorrelated) normal distributions with zero mean and variances 1.0 and 0.01, respectively, and suppose $y_1 = x_1 + x_2$.

Clearly taking 97.5% confidence intervals for x_1 and x_2 and applying interval analysis will give a 95% confidence interval $[-2.47, +2.47]$ for y_1 , but not an optimal one. For example taking a 96% confidence interval for x_1 and a 99% confidence interval for x_2 will give the tighter 95% interval $I = [-2.31, +2.31]$ for y_1 . However the variances of the independent inputs add to 1.01, which is therefore the variance of y_1 , so our approach directly gives $y_1 \in [-1.97, +1.97]$ with 95%.

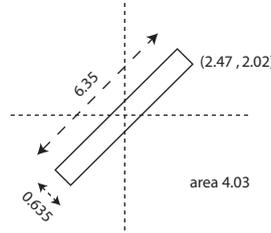


Fig. 1.1. The rectangle J

Now suppose $y_2 = x_1 - x_2$. The square $I \times I$ is a 95% confidence box for $y = (y_1, y_2)$ with area 21.3, although the corresponding square with half-side 1.97 is not. However, in Fig. 1.1, the thin rectangle J aligned at 45 degrees to the axes,

$$J = \{(y_1, y_2) : |y_1 + y_2| \leq 4.49, |y_1 - y_2| \leq 0.45\},$$

is a 95% confidence box and has the much smaller area 4.03.

The optimal 95% ellipsoid E , given by the method of the previous section,

$$E = \{(y_1, y_2) : (y_1 + y_2)^2 + 100(y_1 - y_2)^2 \leq 23.96\}$$

has an area of just under 3.77.

Even in the exactly linear case, if rigorous bounds are sought for the output confidence intervals, then there are benefits to using the approach described here in conjunction with conventional interval analysis, rather than relying upon a naive use of the latter.

The method of this paper can be used to construct a sensible hypothesis for the rigorous approach to verify. The benefit of using the approach described

here to “precondition” the hypothesis for rigorous validation is even more pronounced in the case of correlated inputs or, as we now discuss, non-linear evaluation functions f .

1.4 Non-linear Models

The assumption of linearity of f over the relevant confidence interval is frequently not justified. In other cases, (approximate) linearity is an hypothesis which we wish to use our model to confirm, rather than a matter of blind faith. In these cases, a non-linear model of the effects of evaluating the function f is required.

In the case where the function model is significantly non-linear, the $y_i^{(0)}$ are generally biased estimators for the y_i . In other words, we can no longer assume $E[y_i] = y_i^{(0)}$. Indeed, in the non-linear case the outputs may not even be monotone functions of the inputs over the intervals in question, nor need the $y_i^{(0)}$ be maximum likelihood estimators for the y_i even in the smooth monotone case⁶.

Although maximum likelihood estimators are appropriate for Bayesian inferences such as data assimilation, for many purposes we also require unbiased estimates of quantities associated with the outputs. In particular if we wish to construct confidence intervals (or ellipsoids) from a pdf model for y then we would like to have unbiased estimates for the relevant moment coefficients in order to construct percentile points of the cumulative output pdf.

Such unbiased estimators can be approximated accurately by using AD to propagate higher order Taylor terms.

1.4.1 Higher Order Program Variables

We can redefine the program variables to represent a pyramid of coefficients corresponding to a higher order truncated multivariate Taylor series, so that the variable

$$v_i = (v_i^{(0)}, (v_i^{(j)})_{0 \leq j \leq n}, (v_i^{(jk)})_{0 \leq j \leq k \leq n}, (v_i^{(jkl)})_{0 \leq j \leq k \leq \ell \leq n})$$

represents the truncated Taylor series

$$v_i^{(0)} + \sum_j v_i^{(j)} \zeta_j + \sum_{j \leq k} v_i^{(jk)} \zeta_j \zeta_k + \sum_{j \leq k \leq \ell} v_i^{(jkl)} \zeta_j \zeta_k \zeta_\ell \ .$$

Initializing the input variables to be first order Taylor series as before, and using forward AD for general truncated multivariate Taylor series, we obtain

$$y = y^{(0)} + \sum_j y^{(j)} \zeta_j + \sum_{j \leq k} y^{(jk)} \zeta_j \zeta_k + \sum_{j \leq k \leq \ell} y^{(jkl)} \zeta_j \zeta_k \zeta_\ell + \dots \ .$$

⁶ Although in the smooth monotone case, if p and $q = p/f'$ are the pdf for x and y respectively, then at the maximum likelihood value for y we have $p' = q \cdot f''$.

1.4.2 Taking Expectations

In general by independence of the ζ_i we have

$$E[\zeta_j^p \zeta_k^q \zeta_\ell^r] = E[\zeta_j^p] \cdot E[\zeta_k^q] \cdot E[\zeta_\ell^r] \quad \text{for } j < k < \ell ,$$

and we can usually evaluate terms such as $E[\zeta_j^p]$ from our knowledge of the distribution from which the ζ are drawn. For example for ζ in $N(0, 1)$ we have

$$E[\zeta^{2p}] = (2p - 1)E[\zeta^{2(p-1)}], \quad E[\zeta^{2p-1}] = 0 \quad \text{for } p > 0 ,$$

and corresponding moments can be pre-calculated for other initial distributions.

1.4.3 Unbiased Output Mean

Using these reductions, we see for example that an unbiased estimator of y is

$$\begin{aligned} E[y] = & y^{(0)} + \sum_j y^{(jj)} + S_j y^{(jjj)} + K_j y^{(jjjj)} \\ & + \sum_{j < k} y^{(jjkk)} + \text{5th order terms} , \end{aligned}$$

where S_j is the skew $E[\zeta_j^3]$, and K_j is the kurtosis $E[\zeta_j^4]$. In the usual case where the ζ_j are symmetric, all S_j and all fifth order terms are zero, and the estimator is therefore accurate to order five. If ζ_j is normal then $K_j = 3$.

1.4.4 Unbiased Output Variance

Unbiased estimators for other quantities can be obtained by taking expectations of other variables. For example, an unbiased estimator for the variance of y is

$$\begin{aligned} V[y] &= E [(y - E[y])^2] \\ &= E \left[\left(\sum_j y^{(j)} \zeta_j + y^{(jj)} (\zeta_j^2 - 1) + \sum_{j < k} y^{(jk)} \zeta_j \zeta_k + \sum_{j \leq k \leq \ell} y^{(jkl)} \zeta_j \zeta_k \zeta_\ell + \dots \right)^2 \right] \\ &= \sum_j (y^{(j)})^2 + 2S_j y^{(j)} y^{(jj)} + (K_j - 1)(y^{(jj)})^2 + 2K_j y^{(j)} y^{(jjj)} \\ &\quad + \sum_{j < k} (y^{(jk)})^2 + 2y^{(j)} y^{(jkk)} + 2y^{(jjk)} y^{(k)} + \dots , \end{aligned}$$

where ζ_j has skew S_j and kurtosis K_j . When ζ_j is normal, $S_j = 0$ and $K_j = 3$.

1.4.5 Unbiased Output Covariances

Similarly in the case of more than one output, the covariances correspond to inner products:

$$\begin{aligned}
Cov[y_p, y_q] &= E[(y_p - E[y_p])(y_q - E[y_q])] \\
&= \sum_j y_p^{(j)} y_q^{(j)} + S_j(y_p^{(jj)} y_q^{(j)} + y_p^{(j)} y_q^{(jj)}) \\
&\quad + (K_j - 1)y_p^{(jj)} y_q^{(jj)} + K_j(y_p^{(j)} y_q^{(jjj)} + y_p^{(jjj)} y_q^{(j)}) \\
&\quad + \sum_{j < k} y_p^{(jk)} y_q^{(jk)} + y_p^{(j)} y_q^{(jkk)} + y_p^{(jkk)} y_q^{(j)} + y_p^{(jjk)} y_q^{(k)} + y_p^{(k)} y_q^{(jjk)} + \dots
\end{aligned}$$

1.4.6 Adjoint Expectations

Expectations of the adjoints $\bar{x}_i = \partial y / \partial x_i$ can also be interpreted directly as sensitivities of parameters of the output uncertainties with respect to parameters of the input uncertainties. In the case of a single output y and uncorrelated inputs x_i , and writing $\mu(z)$ and $\sigma(z)$ respectively for the mean and standard deviation of the distribution modelling the uncertainty corresponding to the program variable z , we have for example that

$$\begin{aligned}
E(\bar{x}_i) &= \frac{\partial \mu(y)}{\partial \mu(x_i)}, & E(\zeta_i \bar{x}_i) &= \frac{\partial \mu(y)}{\partial \sigma(x_i)}, \\
\frac{E(y \bar{x}_i) - E(y)E(\bar{x}_i)}{\sigma(y)} &= \frac{\partial \sigma(y)}{\partial \mu(x_i)}, & \frac{E(y \zeta_i \bar{x}_i) - E(y)E(\zeta_i \bar{x}_i)}{\sigma(y)} &= \frac{\partial \sigma(y)}{\partial \sigma(x_i)},
\end{aligned}$$

and so on.

1.5 Implementation with Automatic Differentiation

All the expressions which we have considered, including those for unbiased estimators, can be evaluated automatically by an appropriately enhanced AD tool. It is straightforward to add intrinsics to an AD tool to specify the moments of the ζ_i . Independent variables can be defined to be of type `Normal`, `Student(n)`, `Uniform` etc. A new operator `E` is defined which evaluates the expectation of a program variable. Further intrinsic functions can be defined in terms of `E`, such as

```

Cov (x,y) = E((x-E(x))*(y-E(y)))
K(x) = E((x-E(x))**4)
Bias (x) = term0(x) - E(x)
etc.

```

Reverse AD can be used instead of forward to calculate the multivariate Taylor series coefficients. This makes the accumulation process more efficient, but requires slightly more care in the definitions of the intrinsics: the terms of each series must be assembled into a variant data type, prior to being combined into the expression to which the intrinsic is applied.

Similar remarks apply to the use of interpolation schemes such as those of Bischof et al.⁷[1] or Neidinger [14]. In the case of symmetric distributions the majority of cross-terms do not contribute to expectations, and so need not be computed: in particular, as remarked earlier, a fourth order estimate is automatically accurate to fifth order in this case.

1.5.1 Convergence and Singularities

The order of Taylor series actually required for a given level of accuracy depends on the non-linearity of f . As an extreme example, if

$$y = \exp x^2 \quad \text{with} \quad x = \zeta \in N(0, 1)$$

then the expectation of y is unbounded. Of course, using a truncated distribution for x will give a correct confidence interval for y even in this case: for this example we could restrict ζ to the range $[-5, 5]$, with appropriate changes to the distribution moments. However it is useful that the AD-based model can signal automatically any potentially catastrophic non-linearity of f .

Singularities in f corresponding to plausible values of the input variables can also be dealt with by truncation, but similar difficulties can arise from imaginary poles of f . For example, the smooth function

$$y = \frac{1}{1 + x^2} \quad \text{with} \quad x = \zeta \in N(0, 1)$$

has finite expectation, but the poles at i and $-i$ mean that the Taylor series in ζ for y has only unit radius of convergence. In such cases the simplistic approach of taking expectations term by term and summing produces a divergent series, even though the expectation of the infinite sum may be finite.

This underlines the remark made by Louis Rall during his 2004 talk in Chicago, that AD is not really a local operation. We describe the function f as *almost-linear*, relative to a given input uncertainty pdf and a desired level of confidence for the outputs, when all poles of f are sufficiently distant from the mean, and the radius of convergence about the mean is sufficiently large, to allow the use of a single truncated distribution upon each of the ζ_i .

However in the example of $1/(1 + x^2)$, truncation sufficient to ensure convergence would rule out altogether the possibility of obtaining a 95% confidence interval. Harley Flanders (personal communication) proposes the use of

⁷ Andreas Griewank (personal communication) points out that this would allow a fifth order estimator of n output variable values y to be obtained at a cost of order $n^2/2$ univariate fourth-order Taylor expansions, which is about one third of the cost of computing the full order four tensor.

a *partitioning* approach to provide a general solution for cases like this, and we describe one such approach below.

1.5.2 The Partition Approach to Expectations

The partition approach to evaluating expectations works as follows. Partition the distribution for the ζ_i into boxes $B_j = \prod_i I_{i,j}$ in such a way that only finite boxes have non-negligible probability. By refining if necessary, ensure that the distance from the centre $(c_{i,j})_i$ of each finite box to any pole of f is large compared to the length of the corresponding radius of the box. For each non-negligible box B_j , recalculate the Taylor coefficients for y in terms of powers of $\xi_i = \zeta_i - c_{i,j}$.

Now, using the independence of the ξ_i , and our prior knowledge of the values of the integrals of the ξ_i^k with respect to the pdf for each element of the partition, apply the expectation operator to each box separately and then sum the results. This gives the correct result even for functions which are not almost-linear. Provided standard partitions are used, integrals for the ξ_i^k over $I_{i,j}$ can be pre-calculated and stored in tabular form.

The use of interval methods for optimization requires a similar form of partitioning to that advocated here. Greater accuracy can be obtained by using more Taylor terms or by allowing the partition to be refined.

1.6 Validation of Uncertainty Models

Numerical values of bias terms can be used to determine the validity of the GUM assumptions or to validate the non-linear model being used to construct the confidence intervals for the outputs. This can be done either as an alternative or as a supplement to Monte Carlo Simulation [4, 9]. In the case where a non-linear multistage model is being used, higher order terms of the first stage outputs can be used directly to initialize the second stage inputs.

The generalized output values from an AD-based tool of this type can be used to construct many different confidence intervals, depending not only on the level of confidence required but also on the strength of the assumptions made by the uncertainty model.

1.6.1 Single Output

For a single output variable y , let $I_s(y)$ be the interval defined by

$$I_s(y) = [E[y] - s \cdot \sigma(y), E[y] + s \cdot \sigma(y)], \quad \text{where } \sigma(y) = \sqrt{E[(y - E[y])^2]}.$$

Then $I_{4.48}(y)$ is a 95% confidence interval for y without any assumptions whatsoever on the distribution of y , by Chebyshev's inequality.

However if the kurtosis of y is known to be less than 4.0, an assumption which can be verified by using AD to compute the relevant expectation⁸, then taking $s = 3.0$ gives a 95% confidence interval for y . This is because, setting

$$z = \frac{y - E(y)}{\sigma(y)}$$

and letting p denote the pdf of z , we have that

$$s^4 \cdot P(|z| > s) \leq \int z^4 p(z) dz ,$$

so under the single assumption that the kurtosis is less than 4.0, which we can verify numerically by checking that $E(z^4) \leq 4.0$, setting $s = 3.0$ gives $P(y \notin I_{3.0}(y)) \leq 0.05$, and $s = 4.48$ gives a 99% confidence interval for y .

1.6.2 The Hyperbolic Cosine Transform

More ambitiously, we have

$$\cosh st \cdot P(|z| > s) \leq \int \cosh tz \cdot p(z) dz ,$$

and for z normal the value of the integral is $\exp \frac{1}{2}t^2$. Setting $t = s$ and validating, for a particular value of s and a ‘nearly normal’ z , the single verifiable⁹ assumption that

$$E(\cosh sz) \leq 1.5 \exp \left(\frac{1}{2}s^2 \right)$$

gives

$$P(|z| > s) \leq 3 \exp \left(-\frac{1}{2}s^2 \right) ,$$

and so on.

Estimates of this form are particularly useful for relatively high values of s , corresponding to a requirement for high levels of confidence for which the use of Monte Carlo Methods is problematic. For example setting $s = 4.0$ with the hyperbolic cosine hypothesis gives a 99.9% confidence interval with no further assumptions about the distribution of z .

If more assumptions are made about the moments of y , then tighter confidence intervals can be constructed. If the uncertainty in y is known (or assumed) to be normal then s can be reduced to 1.96 for 95%, 2.58 for 99%, and 3.29 for 99.9%, and similarly for other distributions.

⁸ Naive use of GUM simply assumes *inter alia* that the kurtosis of y is exactly 3.0 if the ζ_i are normal.

⁹ From an implementation point of view, it is important to ensure that the Taylor series developed for \cosh contains no odd terms, rather than being defined in terms of \exp .

1.6.3 Multiple Outputs

Similar observations to the single output case hold for the case of $n > 1$ outputs. If

$$\text{Cov}(y) = E((y - E(y))(y - E(y))^T) = RR^T$$

is the AD-calculated covariance for the outputs, then we can define

$$z = R^{-1}(y - E(y))$$

for the *calculated* values of R . Certainly (for example)

$$s^4 n^2 \cdot P(|z| > s\sqrt{n}) \leq \int z^4 p(z) dz = E((z^2)^2) ,$$

and we can calculate the value of $E((z^2)^2)/n^2$. Under the GUM assumptions, we have $E((z^2)^2)/n^2 \approx 1 + 2/n$, so we can attempt to verify the hypothesis that this value is less than 3.0 (say). In this case, we have

$$P(|z| > s\sqrt{n}) \leq 3.0/s^4 ,$$

and so on.

1.7 Way Ahead

We have shown how to use AD to propagate Taylor series corresponding to uncertainty distributions through functions modelling measurements. A novel point is that the use of AD allows assumptions about the output moments to be validated¹⁰ before a confidence interval at a given level of significance is built. This could be extended to Laurent series. Future work may also include the incorporation of rounding errors and errors arising from the inexact solution of intrinsic equations, using the methods of Iri [3, 7, 10]. A very desirable feature would be the incorporation of interval methods to provide rigorous bounds on the errors in the calculated expectation values. The systematic use of partitioning could also allow automatic validation of control flow changes (such as `if` statements) via partitions of confidence intervals on intermediate values of program variables.

¹⁰ We recommend using AD to validate the assumption that the kurtosis is less than some given value, preferably an hypothesized value significantly larger than the calculated one. This is different from using the calculated value to construct the confidence interval. It is the former procedure which we are recommending here.

References

1. Christian Bischof *et al*, 1993, Structured Second- and Higher-Order Derivatives through Univariate Taylor Series, *Optimization Methods and Software*, **2**, 211–232.
2. Bruce Christianson, 1992, Reverse Accumulation and Accurate Rounding Error Estimates for Taylor Series Coefficients, *Optimization Methods and Software*, **1**(1), 81–94.
3. Bruce Christianson, 1998, Reverse Accumulation and Implicit Functions, *Optimization Methods and Software*, **9**(4), 307–322.
4. M.G. Cox, M.P. Dainton *and* P.M.Harris, 2001, Uncertainty and Statistical Modelling, Software Support for Metrology Best Practice Guide No. 6, HMSO, ISSN 1471-4124.
5. M.G. Cox *and* P.M.Harris, 2004, Uncertainty Evaluation, Technical Report, National Physical Laboratory, Teddington UK.
6. Andreas Griewank, David Juedes *and* Jean Utke, 1996, Algorithm 755; ADOL-C: A package for the automatic differentiation of algorithms written in C/C++, *ACM TOMS*, **22**(2), 133–167.
7. Masao Iri, 1991, History of Automatic Differentiation and Rounding Error Estimation, pp 3–16 *in* Andreas Griewank *and* George Corliss (Editors), 1991, *Automatic Differentiation of Algorithms*, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania.
8. International Organisation for Standardisation, 1995, Guide to the Expression of Uncertainty in Measurement, 2nd edition, Geneva, ISBN 9267-10188-9.
9. International Organisation for Standardisation, 2004, Guide to the Expression of Uncertainty in Measurement, Supplement 1: Numerical Methods for the Propagation of Probability Distributions, Draft Technical Report, Joint Committee for Guides in Metrology.
10. Keiko Kabaya *and* Masao Iri, 1987, Sum of Uniformly Distributed Random Variables and a Family of Nonanalytic C^∞ -Functions, *Japan Journal of Applied Mathematics*, **4**(1), 1–22.
11. Gershon Kedem, 1980, Automatic Differentiation of Computer Programs, *ACM TOMS*, **6**, 150–165.
12. Koichi Kubota *and* Masao Iri, 1991, Estimates of Rounding Errors with Fast Automatic Differentiation and Interval Analysis, *Journal of Information Processing*, **14**(4), 508–515.
13. Ignacio Lira, 2002, *Evaluating the Measurement Uncertainty: Fundamentals and Practical Guidance*, Institute of Physics Publishing, Bristol.
14. Richard Neidinger, 2005, Directions for Computing Truncated Multivariate Taylor Series, *Math Comp*, **74**(249), 321–340. See www.ams.org/mcom/2005-74-249/home.html