940722

Uncertainty in Accident Reconstruction Calculations

Raymond M. Brach
University of Notre Dame

ABSTRACT

The problem of determining the uncertainty in the result of a formula evaluation is addressed. The origin of the uncertainty is the presence of variations in the input variables. Three popular techniques are discussed in the context of accident reconstruction. The first establishes upper and lower bounds through calculation of the largest and smallest possible values of the quantity being estimated for all combinations of the input variables. The second method uses differential calculus and places variations of the variables into a delta equation derived from the mathematical formula. The last method covers cases where statistical information about the input data is known. Approximate means and variances are developed for linear and nonlinear formulas. Examples are given for all of the methods such as calculation of speed from skid distance and calculation of stopping distance including perception-decision-reaction (PDR) time.

INTRODUCTION

Because of their intended purpose many reconstruction calculations find their way into courtroom testimony. According to Bernstein (1993) there has been recent legal activity pertaining to the screening of technical data and expert testimony. Federal Rules of Evidence, according to Bernstein, now include the following:

1. the court should determine whether the theory or technique in question can be (or has been) tested
2. peer review is an important consideration
3. the known or potential rate of error of the technique should be determined, as should be the existence and maintenance of standards controlling the technique's operation.

This paper addresses some of the methods useful in addressing the item 3 in the context of vehicle accident reconstruction. Although some theory of uncertainty is presented, coverage is primarily by example to illustrate some common ways of calculating and expressing uncertainty.

Following some definitions, three techniques are discussed. These are the establishment of upper and lower bounds on a calculated quantity, the use of differential calculus to provide a way to determine uncertainties in calculations and lastly, uncertainty in a statistical context.

SOME DEFINITIONS

Because the topic of uncertainty in measurements and calculations arises in many fields, many terms are used to describe the concept of "uncertainty". For example, books on the topic of *uncertainty* use synonyms and modifiers such as variation, error, disturbance, discrepancy, fluctuation, etc. Some definitions and descriptions are now given, not with the intention of answering any questions in terminology, but rather to make sure that the readers of this paper understand the terms being used.

The context here is that of using a *formula* or equation to *estimate* an accident *variable* which depends on the values of constants (whose values are known with "certainty") and other *variables* representing physical quantities which can possess significant variations. In common mathematical terms the estimate is the dependent variable, y, which depends on one or more independent variables (physical quantities), u, v, . . . , w. Often, these independent variables will be referred to simply as the variables. In this paper, the term *variation* is arbitrarily associated with the independent variables and the term *uncertainty* is associated with the estimate, y. Special names given to specific values of some of the variables such as u will be discussed including the minimum, u_{mn}, maximum u_{mx} and a representative or nominal value, U. Variations in the variables will be denoted by δu, δv, . . . , δw, leading to a corresponding uncertainty, δy. The origin or cause of the variations is broad and can come from: repeated measurements, unmeasured physical quantities represented inexactly by "typical" values, quantities known only with a limited degree of precision, etc.

UPPER AND LOWER BOUNDS ON ESTIMATES

One of the simplest ways of quantifying uncertainty is to establish upper and lower bounds on the dependent variable caused by variations in the independent variables. First, those quantities in the equation which possess a significant degree of variation are identified as variables. Then reasonable ranges of each variable are determined. Finally the lowest and highest values of the dependent variable for all possible combinations of the values of the independent variables are computed.

For example, one of the most commonly used formulas in accident reconstruction is one which uses a measured length of skid marks to provide an estimate of the speed of a vehicle at the instant the wheels lock. The formula for the initial speed v_I preceding a locked-wheel skid over a distance d ending at a speed of v_F, is

$$v_I = (v_F^2 + 2\ f\ g\ d)^{1/2} \qquad (1)$$

Under typical circumstances, the quantities f and d, the friction factor and skid mark length, are known with less than perfect certainty and are chosen as the variables here. v_F might also be a variable, but is arbitrarily assumed here to be zero. The remaining quantity, the acceleration of gravity, g, is known accurately and is taken to be a constant. Suppose that d and f are known with uncertainty such that $d_{mn} \leq d \leq d_{mx}$ and that $f_{mn} \leq f \leq f_{mx}$. A corresponding upper and lower bound on the estimate of v_I is

$$(2\ f_{mn}\ g\ d_{mn})^{1/2} \leq v_I \leq (2\ f_{mx}\ g\ d_{mx})^{1/2} \qquad (2)$$

The uncertainty δv_I is taken to be half of the difference of the upper and lower bounds so that $V_I \pm \delta v_I$ is the way the estimate of speed at wheel lock is written to show uncertainty. For example if $f_{mn} = 0.6$, $f_{mx} = 0.8$, $d_{mn} = 32.0$ m and $d_{mx} = 34.0$ m then 19.4 m/s $\leq v_I \leq 23.1$ m/s. Then $\delta f = 0.1$, $\delta d = 1$ m and the value of V_I with uncertainty is 21.3 ± 1.9. (According to common practice the number of significant figures to the right of the decimal point is chosen to reflect the precision of the input data.)

This example illustrates what is probably the simplest and most versatile method for determining uncertainty. It applies to any formula no matter how complex and is easy to carry out. It is even possible to use this with computer simulations using multiple runs with different input. Care must be used when the formula involves differences and division. For example, the lower limit of y = (a-b)/c is obtained by using the lower limit of a and the upper limits of b and c. Negative numbers can also be tricky.

A drawback of this (and the next) method is that the statistical nature of the variations is not explicitly taken into account and so the likelihood or probability of reaching either of the limits cannot be assessed. Attributing the upper and lower bounds to a specific percentage of a population should not be done, or be done with due caution. Statistical conclusions should follow the use of statistical methods and always be based on statistical data.

DIFFERENTIAL VARIATIONS BASED ON THE MATHEMATICAL FORMULA

Another common method of estimating uncertainty, frequently referred to as a form of *error analysis* is covered in many laboratory courses taken in science and engineering. For example, see texts such as Beers (1957) and Taylor (1982). The method uses differential calculus to relate the quantity being calculated, y, to the dependent variables u, v, . . . , w. In general

$$y = f(u, v, \ldots, w) \quad (3)$$

From calculus, the differential of y can be found using the chain rule as

$$dy = \left(\frac{\partial y}{\partial u}\right) du + \left(\frac{\partial y}{\partial v}\right) dv + \ldots + \left(\frac{\partial y}{\partial w}\right) dw \quad (4)$$

With the nominal or reference values of u, v, . . ., w given by U, V, . . . , W, a formula for uncertainty is found by replacing the variable differentials by variations such that

$$\delta y = \left(\frac{\partial y}{\partial u}\right)_{U,V,W} \delta u + \left(\frac{\partial y}{\partial v}\right)_{U,V,W} \delta v + \ldots + \left(\frac{\partial y}{\partial w}\right)_{U,V,W} \delta w \quad (5)$$

Equation 5 is an approximation that amounts to a linearization of the function y(u, v, . . . , w) around its value y(U, V, . . . , W). Note that the derivatives are evaluated at the reference or nominal values.

In the example of Eq 1, $y = v_I$ and there are 2 variables, $u = f$ and $v = d$. After using Eq 5 and dividing by the nominal value V_I, the relative variation of v_I is

$$\frac{\delta v_I}{V_I} = ½ \left[\left(\frac{\delta f}{F}\right) + \left(\frac{\delta d}{D}\right) \right] \quad (6)$$

The nominal values for d and f are the mid or averages of d_{mn} and d_{mx} and of f_{mn} and f_{mx}. Finally

$$\delta v_I = 10.64 \left[\left(\frac{0.1}{0.7}\right) + \left(\frac{1.0}{33.0}\right) \right] = 1.8 \text{ m/s}$$

It is clear that in this example, the variations in friction have a considerably greater effect than the variations in the distance measurements. This would generally be true especially when the friction factor is small, such as under icy conditions.

Note that in the previous section, v_I was bounded by 23.1 and 19.4; subtraction gives a $\delta v_I = 1.9$, similar to above. Recall that the upper and lower bounds found in the previous section do not involve any approximation whereas the derivation of Eq 5 involves replacement of the (infinitesimal) differentials by (finite) variations. Further recall from calculus that this process is equivalent to expanding y in a Taylor series about U, V, . . ., W, and dropping all higher order derivatives. So the fact that the two example uncertainties are identical indicates that the approximation of Eq 5 is a good one for the uncertainty of Eq 2 in the region of the given nominal values.

In general, if

$$y = a\, u^p v^q \quad (7)$$

where a, p, and q are constants, then

$$\delta y = p \frac{Y}{U} \delta u + q \frac{Y}{V} \delta v \quad (8)$$

For equations of the form of Eq 7 (such as Eq 1), the uncertainty depends not only on each variation but also is influenced by the ratios, Y/U and Y/V. Where at least one ratio is large, the uncertainty is large. For the skid distance equation, the uncertainty generally increases as the nominal coefficient of friction becomes smaller. In fact, if the nominal friction factor in the example was much lower, say 0.2, then the uncertainty contributed by the distance term would be negligible.

STATISTICAL UNCERTAINTY

In the skid distance example with bounds on the velocity estimate given by Eq 2, a range of values of f, $f_{mn} = 0.6$ to $f_{mx} = 0.8$, was chosen. If measurements of the friction factor are not made at an accident scene, a typical range such as this is often used, chosen from experience and/or from published data. This particular range would be a reasonable estimate for well-travelled dry asphalt pavement at moderate to high speeds, for example. But the exact meaning of f_{mn} and f_{mx} is subject to ambiguity however. Are these bounds that are never exceeded? Are they values exceeded only 1 or 2% of the time? Furthermore, is an intermediate value such as 0.66 as likely as that of 0.71; is 0.60 as likely as 0.73, or 0.75? Or, suppose that a single value of $f = 0.72$ was measured at an accident scene. Then what variation should be used? If the likelihood of the values or ranges of values of f are to be taken into account, then statistical theory must be used. The problem then is to find the distribution of the uncertainty for given distributions of the independent variables, now considered to be random variables.

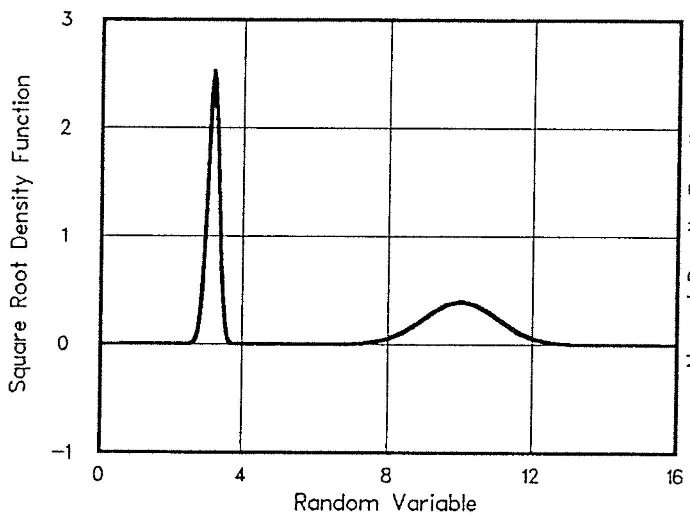

Figure 1. Probability density functions for a normally distributed variable with a mean of 10 and a variance of 1 (on right) and the square root of a normal distribution (on left) with approximate mean and variance of 3.16 and 0.158, respectively

There is a well known theorem from statistical theory of normal or Gaussian distributions that can be very helpful in relating variations to uncertainty. If a random variable has a normal distribution, say $u = N(\mu_u, \sigma_u^2)$, where μ_u is the mean and σ_u^2 is the variance, and a quantity, x, depends linearly on u,

$$x = a\,u + b \tag{9}$$

where a and b are constants, then x is also normally distributed, $x = N(\mu_x, \sigma_x^2)$, where $\mu_x = a\,\mu_u + b$ and $\sigma_x^2 = a^2\,\sigma_u^2$. (See Hald, 1960, for additional information.) For example, for a constant speed, v, the distance travelled is

$$x = v\,t \tag{10}$$

where t is time. If a set of measurements of t are normally distributed then the distance calculated from Eq 10 is normally distributed, according to the theorem. Unfortunately, linear relationships seldom play a major role in accident reconstruction. For some special cases, some nonlinear relationships can be handled rigorously. One of these, a form of Eq 1, is discussed below. But a more general, approximate theory is also available; this is also discussed followed by an example of a stopping distance calculation.

Distribution of a Square Root Function According to Hald (1960), if an equation relating two random variables has the form

$$x = a\,\sqrt{u}, \qquad a > 0 \tag{11}$$

and $u = N(\mu_u, \sigma_u^2)$ then x has a probability density function

$$f(x) = \left(\frac{2}{\pi}\right)^{1/2} \frac{x}{a\sigma_u}\, e^{-(x^2/a - \mu_u)^2/2\sigma_u^2} \tag{12}$$

where the mean and variance of x are given by $\mu_x \approx a\,\sqrt{\mu_u}$ and $\sigma_x \approx a\sigma_u/2\sqrt{\mu_u}$. Equation 12 is not the same as a normal distribution but has a similar shape. Figure 1 shows a typical pair of density functions corresponding to Eq 11 and 12. Equation 1 has the form of Eq 11 when $v_f = 0$ and if only f has a significant variation (i.e., the distance has negligible variation). Figure 1 illustrates that even if f has a relatively broad statistical variation (curve on the right) the uncertainty in the velocity (distribution on the left) can be relatively small.

Consider such an example with $v_F = 0$, a known skid length and with the friction factor alone as a random variable. For compatibility with the previous examples, values of d = 33 m and a mean value of f, $\mu_f = 0.7$ (mid way between 0.6 and 0.8), are chosen. A variance $\sigma_f^2 = 5.0 \times 10^{-3}$ is used (typical of actual experimental values, presented shortly). Then the initial velocity has an approximate mean and standard deviation (square root of the variance) of $\mu_v = 21.3$ m/s and $\sigma_v = 1.08$ m/s. Under the further assumption that f is approximately normally distributed, various statements about the initial speed estimate can be made. Approximately 90% of a normal variable has values between $\mu \pm 1.645\sigma$, so for μ_v and σ_v as calculated, the skid distance for the given pavement frictional characteristics would indicate an initial speed of 21.3 ± 1.8 m/s about 90% of the time. Note that despite that δv_1 from example 2 above and $1.645\sigma_v$ here both equal 1.8 m/s, only the latter has a specific statistical meaning.

Distribution of More General Functions For nonlinear mathematical formulas or equations which relate random variables, approximate methods must be used to find the statistical uncertainty. The mean and variance of a variable y that depends on random variables u, v, . . . , w such that

$$y = y(u, v, \ldots, w) \tag{13}$$

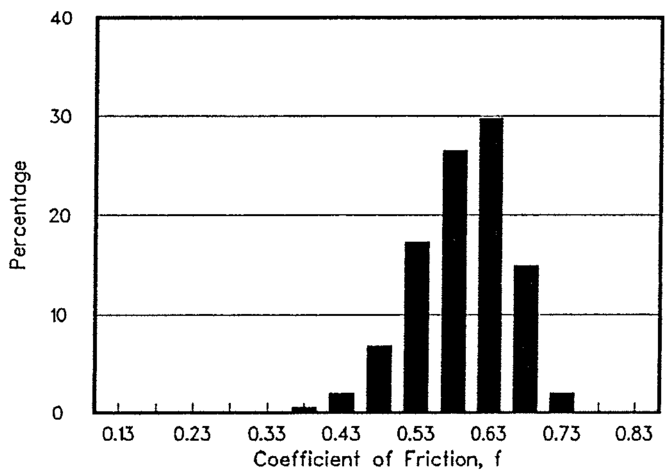

Figure 2. Measurements of the coefficient of friction over the full length of a new, 930 km asphalt road in Holland; from a TRRL Report (1972). The average is 0.593 and the standard deviation is 0.065.

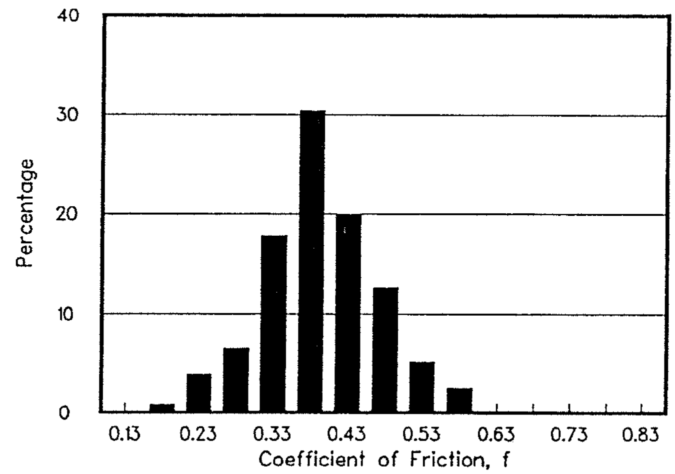

Figure 3. Distribution of skid numbers for 230 test sections of two-lane Kentucky roads for a wide range of annual average daily traffic; Rizenbergs, et al (1977). The average is 0.393 and the standard deviation is 0.079.

can be found by expanding y in a Taylor series, dropping terms with derivatives higher than the first and using the relationship for linearly related normal variables mentioned earlier. The procedure is given in detail in Taylor (1982) and Beers (1957) and is not repeated here. In summary, the mean of y is

$$\mu_y \approx y(\mu_u, \mu_v, \ldots, \mu_w) \tag{14}$$

That is, the mean of y is found by evaluating Eq 13 with the mean values for the variables. The variance of y is

$$\sigma_y^2 \approx \left(\frac{\partial y}{\partial u}\right)^2 \sigma_u^2 + \left(\frac{\partial y}{\partial v}\right)^2 \sigma_v^2 + \ldots + \left(\frac{\partial y}{\partial w}\right)^2 \sigma_w^2 \tag{15}$$

where the derivatives are evaluated at the mean values of the variables.

Example Figures 2 and 3 show the results of measurements of pavement-tire friction coefficient on the roads indicated. The distribution of values in both cases is approximately bell-shaped which indicates that an assumption of normality can be made for those pavements. Normality of tire pavement frictional variations is used in the following example of determining the statistical properties (mean and variance from Eq 14 and 15) of stopping distance. Stopping distance is the total distance, d_T, travelled during the perception-decision-reaction time, t_P, and the skidding distance, d_S, during a panic, locked-wheel skid to rest. The total distance is

$$d_T = (2gd_S)^{1/2} \sqrt{f}\, t_P + d_S \tag{16}$$

It is assumed here that the perception-decision-reaction time, and the friction factor, are normally distributed random variables. The data for this example is summarized in Table 1.

Equation 14 gives an average stopping distance of

$$\mu_T = a\, \mu_v\, \mu_P + d_S = (2gd_S)^{1/2} \sqrt{\mu_f}\, \mu_P + d_S$$

$$= (25.45)\,(0.7)\,(1.50) + 33 = 64.9 \text{ m}$$

Equation 15 gives

$$\sigma_T^2 = \mu_P^2\, \sigma_v^2 + \mu_v^2\, \sigma_P^2 = 168.4$$

or

$$\sigma_T = 13.0 \text{ m}$$

If an assumption is made that the stopping distance is normally distributed, then 90% of the population is contained within $\mu_T \pm 1.645\, \sigma_T$, that is, the stopping distance is 64.9 ± 21.4 m about 90% of the time. Another way of stating this is that, based on the data

Table 1
Data for the Stopping Distance Example

Variable	Type	Value or Mean & Variance
g	constant	$9.81\ m/s^2$
d_S	constant	$33\ m$
f	$N(\mu_f, \sigma_f^2)$	$\mu_f = 0.7$
		$\sigma_f^2 = 5.0 \times 10^{-3}$
t_P	$N(\mu_P, \sigma_P^2)$	$\mu_P = 1.50\ s$ *
		$\sigma_P^2 = 0.36$ *

* these are representative values from Olson (1989); also see Sens, et al (1989)

given in Table 1 and with 90% confidence, the stopping distance is between 43.5 and 86.3 m.

An assumption is made in deriving the method given by Eq 13, 14 and 15 and should be understood when applying it. The assumption is that the variations of u, v, . . . , w are statistically independent. Consider this in the context of Eq 1. Suppose the distance d is measured at an accident scene by pacing-off the distance and multiplying the number of paces by the average of the person's step length. Suppose further, that a vehicle is brought up to a known speed, its wheels are locked and the skid distance measured to estimate the friction factor (by calculation). If the skid length for the skid test is measured by the same person using the same method, then variation in f is tied in or related to variation in d, that is, they are not independent. If the relationship in the variations in the variables is known, the independence assumption can be overcome. This is not dealt with here and independence of the variations is assumed.

DISCUSSION AND CONCLUSIONS

Questioning the accuracy and reliability of estimates for accident reconstructions in a legal environment is certainly legitimate. If courts become more demanding (and they may very well in the future) techniques such as above will become more commonplace. Obviously, the more accurate the input, the more realistic and reliable the results. The weight of testimony should be in proportion to its accuracy. Critical cases should not be decided on results with high uncertainty.

On the positive side, calculations of uncertainties can make reconstructions more effective when accurate input information is available. If calculating uncertainties is to become more commonplace, it will be necessary to determine and record the variability of measurements at the time the measurements are made. If a skid length fades over a distance of 1 m, then it should be measured as a length ± ½ m variation. In other words, the concepts here should be introduced to process of accident investigation.

All three methods covered above are useful to estimate uncertainty. The second method using differential variations is more sensitive to the mathematical form of the equation being used and is more illustrative in comparing relative uncertainty due to the different variables. Its use should be limited to situations where the variations are relatively small because of the linearization used in its derivation. Computing upper and lower bounds is the simplest and most general of the three methods to apply but provides the least information. Since the likelihood of simultaneously reaching the extreme values of the variables is not taken into account, the results can be unrealistic.

The last, using mathematical statistics, provides the most information about uncertainty and requires the most input information, namely, quantitative statistical descriptions (distributions) of the independent variables. Being an analytical method, it is necessary to use approximate distributions for nonlinear formulas. It can work well for simple to moderately complex reconstruction equations, but becomes impractical for complex reconstruction problems.

The methods covered in this paper are not the only ways of estimating uncertainty. A method currently receiving attention is the use of Monte Carlo methods. Roughly speaking, the Monte Carlo method is a brute force randomized simulation on a computer of a mathematical model using appropriate statistical distributions for each of the variables. It can be quite sophisticated and is capable of handling complex models as well as correlations among the problem variables.

ACKNOWLEDGEMENT

The author expresses appreciation to Rolly Kinney, Kinney Engineering, Corvallis, OR, for his helpful comments and suggestions which led to several improvements in this paper.

REFERENCES

Beers, Y., (1957), *Introduction to the Theory of Error*, Addison-Wesley, Reading, MA

Bernstein, D., (1993), "Hauling Junk Science Out of the Courtroom", *Wall Street Journal*, July 13

Hald, (1960), *Statistical Theory with Engineering Applications*, John Wiley & Sons, New York, NY

Olson, P. L., (1989), "Driver Perception Response Time", Paper 890731, SAE, Warrendale, PA 15096-0001

Rizenbergs, R. L., J. L. Burchett and L. A. Warren (1977), "Relation of Accidents and Pavement Friction on Rural, Two-Lane Roads", Transportation Research Record 633, *Pavement Surface Properties and Performance*, National Academy of Science, Washington, DC

Sens, M. J., et al, (1989), "Perception/Reaction Time Values for Accident Reconstruction", Paper 890732, SAE, Warrendale, PA 15096-0001

Taylor, J. R., (1982), *An Introduction to Error Analysis*, University Science Books, Oxford University Press, Mill Valley, CA

TRRL Supplementary Rpt 24UC, (1972), Proceedings of the Symposium on Skid-Resistant Roads, Crowthorne, UK