

# Determining Parameter Limits in Traffic Crash Reconstruction Using Limited Data

Jeremy S. Daily\*

October 19, 2006

## Abstract

Quantities used in crash reconstruction often have inherent variation and a single value is not appropriate. A common way to overcome this deficiency is to use a range. This paper will present a technique, based on sampling statistics, to determine a range in a logical and mathematically consistent fashion.

If the parameter in question is normally distributed, then the Student-t and  $\chi^2$  (chi-squared) distributions from sampling statistics provide descriptions of the mean and variance respectively. The number of samples determine the characteristics of both the Student-t and  $\chi^2$  distributions. In a real world application, both the mean and variance are unknown so simplifying assumptions found in statistics texts do not apply. In fact, compounding all the uncertainties leads to a larger precision interval than desired. Therefore, a stochastic equation is presented that requires numerical integration to obtain the final distribution of the parameter. The precision interval (range) obtained from this distribution contains the uncertainty associated with limited samples. As the number of samples increase, the integrated distribution approaches a normal distribution defined by the sample mean and sample standard deviation. A numerical example demonstrates convergence. The results may give an unacceptably large range in which additional samples are required. A traffic crash reconstruction related example is included.

---

\*Assistant Professor of Mechanical Engineering, The University of Tulsa, 600 S. College Ave., Tulsa OK 74104-3189, PH: (918) 631-3056, e-mail: jeremy-daily@utulsa.edu

**Keywords** Statistics, probability, traffic crash reconstruction, accident analysis

# 1 Introduction

## 1.1 Motivation

Many technical processes and problems have variation associated with them. This is especially true in the field of traffic crash reconstruction or other forensic investigations where many of the data are obtained through measurements that are subject to the ability of the investigator. For example, the basic premise of traffic crash reconstruction is to give an assessment of the state of the vehicles or parties involved before the crash based on measurable and anecdotal evidence from the crash scene. These measurements, whether taken at the scene or some date later, are only representative of the true or correct value for a particular crash event. Some error<sup>1</sup> exists with any measurement about the correct values so precision intervals, or ranges are commonly used to represent the uncertainty. The goal of this paper is to provide a mathematically consistent technique for determining a precision interval based on limited data.

For example, the drag factor of a certain vehicle/road system may be desired during the investigation of a crash. To determine drag factor, an investigator uses an accelerometer in an exemplar vehicle on the same surface of the crash to obtain the following measurements: 0.763, 0.720, 0.751, and 0.743. Notice that these measurements have exceeded the 5% spread which has commonly been taught as a criteria for “good” measurements in crash reconstruction (Rivers, 1997). The arithmetic mean of the samples is  $\bar{x} = 0.744$ ; however, obtaining another measurement may change the value of the estimated mean. The the estimated standard deviation (a measure of spread) for the above data is  $s = 0.0181$ . In a similar manner, the estimated standard deviation may change with subsequent measurements, or, the actual value of the drag factor for the actual crash under investigation may be slightly lower or higher than the current measurements indicate.

If only the mean value is of interest, a well established uncertainty analysis can be performed

---

<sup>1</sup>Error is defined as the difference between the measured value of the variable and the true value of the variable.

(Dieck, 2002). However, in a forensic context, all probable values must be considered. Therefore, the range must be estimated from the mean and standard deviation— both of which are unknown. For this situation, introductory texts on statistics do not address the procedure for estimating a precision interval when both parameters are unknown. Simply combining the 95% confidence interval for the mean with the 95% confidence interval for the standard deviation will lead to an interval that is greater than 95%. Therefore, a Monte Carlo based approach is employed in this paper.

A so-called good measurement must repeatably and accurately represent the phenomenon being measured at reasonable cost. The reader should always ensure the data used in an investigation are being applied correctly and no blunders are present in an analysis. The techniques presented herein assume the data are valid and are being applied according to sound mathematical and physical principles (Daily et al., 2006).

Statistical analysis does not turn erroneous data into valid data. Furthermore, only the data gathered are being used. This means any prior knowledge is ignored. Other approaches exist, such as Bayesian analysis, that can include prior knowledge in an analysis as introduced by (Roberson and Vignaux, 1995; Marks, 2002; Aitken and Taroni, 2004).

## **1.2 Background**

The technical aspect of this paper requires the reader to have a working understanding of statistics as outlined in introductory texts (Ayyub and McCuen, 1997). However, a brief explanation of the statistical terms will be presented in this section.

The goal of this paper is to demonstrate a technique of obtaining estimates for the lower and upper bounds of a range to use in subsequent deterministic calculations. The word *deterministic* means the values in an expression are single valued. In other words, there is only one value for each variable. However, in a *stochastic* problem, one or more variables can have multiple values, that is, the variable is random. When this occurs, one must employ non-deterministic methods to perform the calculations.

A stochastic variable can be classified into a few main groups. In general, a variable can be either discrete or continuous. A discrete quantity is something that can be counted, like the number of vehicles traveling through an intersection, and a continuous quantity is something one measures; for example, length or acceleration. This discussion will deal with continuous variables and using measurements to determine a range of possible values.

There are two types of uncertainty associated with a variable as outlined by (Oberkampf et al., 2001). The first is type is called *aleatory* uncertainty which describes the inherent variability associated with a particular quantity. This can be thought of as the noise of the system. The second type of uncertainty is called *epistemic* uncertainty. This uncertainty is associated with lack of information. As the amount of information about a particular variable increases, the epistemic uncertainty decreases. When estimating parameters with limited data, both aleatory and epistemic uncertainty exist. Therefore, the approach presented herein represents both the aleatory and epistemic uncertainties.

### **1.3 Random Variables**

Examples of random variables in traffic crash reconstruction include drag factor, crush stiffness values, pedestrian walking speeds, perception-response time, and distance measurements. Distance measurements are random,<sup>2</sup> not because the length is always changing, but because of different measuring techniques and the human error associated in detecting beginning and ending points. Clear examples of measurement variation common in traffic crash investigation are demonstrated by (Bartlett et al., 2002). Even if an event was recorded and a desired parameter was measured, there exists uncertainty inherent to the measurement process.

Of the many ways to represent uncertainty, probability functions are widely adopted. If one has complete knowledge of a continuous probability function, then probabilities can be assigned to certain intervals. Probability functions must adhere to the following axioms of probability (Kreyszig, 1970):

---

<sup>2</sup>The words random and stochastic are interchangeable in this paper.

1. The probability of an event,  $E$ , in a sample space,  $S$  is between 0 and 1. This is written mathematically as:

$$0 \leq P(E) \leq 1 \quad (1)$$

2. The probability of the entire sample space is unity. Mathematically:

$$P(S) = 1 \quad (2)$$

3. If events in the sample space are mutually exclusive, then the probability of the union of all events are the sum of the individual probabilities. Expressed symbolically as:

$$P(E_1 \cup E_2 \cup \dots) = P(E_1) + P(E_2) + \dots \quad (3)$$

In essence, axiom #3 allows for the determination of probability by integrating the probability density function.

The previous three axioms refer to an event. In general, this event can be anything, however, this discussion is limited to measurements. Specifically, a measurement event is represented by an equality:

$$E \rightarrow X = a \quad (4)$$

where  $E$  is the event,  $X$  is the random variable, and  $a$  is an observed value.<sup>3</sup> The probability of such an event is written as  $P(X = a)$ .

When dealing with real systems and measurements, obtaining an exact equality is impossible. Therefore, the event must be translated into an inequality:

$$E \rightarrow a \leq X \leq b \quad (5)$$

---

<sup>3</sup>Mathematical convention uses capital letters to represent random variables. Lowercase letters correspond to specific observations of a random variable.

where  $a$  is a lower bound and  $b$  is an upper bound for the variable  $X$ . As the difference between  $a$  and  $b$  becomes small, the description of a variable transitions from a discrete random variable to a continuous random variable. Distribution functions are used to describe continuous random variables.

### 1.3.1 Distribution Functions

A *cumulative distribution function* (CDF) represents the total probability of the random variable having a value lower than some value  $x$ . The CDF is commonly represented by a capital letter and is computed using integration:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(s) ds \quad (6)$$

where  $F(x)$  is the CDF,  $X$  is the random variable,  $x$  is a particular value of the random variable, and  $s$  is a dummy variable of integration. The integrand is always positive, so the CDF is always increasing. The CDF always starts at zero on the left and increases to unity on the right of the graph. The integrand,  $f(s)$ , is known as the *probability density function* or *probability distribution function* (PDF).

The PDF is the derivative of the CDF and commonly is represented by a lowercase letter:

$$f(x) = \frac{dF(x)}{dx} \quad (7)$$

Once the PDF is known, the probability of any interval can be calculated using integration:

$$P(a \leq x \leq b) = \int_a^b f(x) dx \quad (8)$$

Often, as is the case in this paper, the evaluation of the integral in Eq. (8) must be performed numerically. The Monte Carlo integration scheme (Weisstein, 2006) is widely popular as well as other advanced statistical sampling schemes (Efron, 1982).

## 1.4 Organization

This paper is organized in the following fashion. A description of the normal distribution along with its assumptions is outlined. Next, the sampling distributions used with limited data are described. This includes the Student-t and the  $\chi^2$  (chi-squared) distributions. The next section shows a technique of combining the sampling distributions using Monte Carlo techniques to get an overall distribution and subsequent precision interval. This precision interval is the desired range, which is the main contribution of this paper. Finally, a numerical example showing the nature of convergence is presented.

## 2 The Normal Distribution

### 2.1 The Most Likely Distribution

Many times an analyst will have to assume a shape for a distribution. The normal distribution is the most likely distribution for an unknown naturally occurring quantity. Moreover, the techniques presented in this paper are robust, meaning that if the underlying distribution is slightly non-normal, then the estimates provided are still good. Gross deviation from normality does render the results meaningful and caution must be exercised to ensure the quantity under analysis is normal or near normal. This can be done visually with probability paper or a more advanced statistical test (Sheskin, 2000).

### 2.2 Definition

The probability density function (PDF) of a normal (Gaussian) distribution is given as:

$$f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[ -\frac{(x-\mu)^2}{2\sigma^2} \right] \quad (9)$$

where  $\mu$  is the mean of the distribution and  $\sigma$  is the standard deviation of the distribution. When a normal distribution is used to describe the population of all possible values of the variable  $x$ , then  $\mu$  is known as the population mean and  $\sigma$  is the population standard deviation. The quantity  $\sigma^2$  is the population variance.

A graph of the PDF of a normal distribution is shown in Fig. 1a and the CDF is shown in Fig. 1b. The precision interval is shown graphically where any random sample drawn from the distribution has a 95% chance of being within the interval indicated. Obviously, if the precision interval was made larger, then the probability of having a value within the defined range increases. Likewise, tightening the precision interval will decrease the probability of a sample being contained within the defined range.

No closed form equation exists for the curve of the CDF of a normal distribution function. Most texts on statistics contain a table with the CDF values. These values correspond to the standard normal distribution where  $\mu = 0$  and  $\sigma = 1$ .

## 2.3 Central Limit Theorem

The central limit theorem says, in essence, that if multiple random variables are added, then their sum follows a normal distribution. There are numerous independent and random factors that affect a physically measurable system. Error from multiple sources is modeled using addition, which means the total error associated with a measurement is likely to follow a normal distribution due to the central limit theorem. The central limit theorem requires a sufficiently large number of sources for the proof, but results converge quickly (near 10 summations) to the Gaussian distribution.

Aside from using the central limit theorem as mathematical justification for the normal distribution, anecdotal evidence strongly suggests many physical and natural phenomena follow a classic “bell” curve which is well represented by the normal distribution. That being said, the normality assumption of the underlying data should be examined for each variable under analysis.

### 3 Sampling Statistics

Assume a quantity follows a normal distribution and  $n$  samples are taken from that population. For example, an investigator conducts four drag factor tests ( $n = 4$ ) and it is assumed that drag factor follows a normal distribution. This normal distribution has an unknown mean and an unknown variance. From the  $n$  samples, one can obtain estimates of those parameters. Of the different estimators, the most common used are the arithmetic sample mean:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (10)$$

and the sample variance:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (11)$$

The sample standard deviation is the square root of the variance ( $s = \sqrt{s^2}$ ). These estimators are purely a function of the gathered data. However, when making assertions about the quality of these estimates, the sample size is important.

#### 3.1 Statistics of the mean

As the number of samples increases, the estimated mean,  $\bar{x}$ , converges to the population mean. However, with a small number of samples, the estimated mean may be different than the population mean. A precision interval can be constructed around the sample mean that will include the actual population mean. The construction of this precision interval is based on the Student-t distribution and the standard error ( $serr$ ). The formula for the standard error is

$$serr = \frac{s}{\sqrt{n}} \quad (12)$$

The precision interval for the population mean utilized the Student-t distribution along with the sample size according to the following equation:

$$\mu = \bar{x} \pm t_{\alpha,n} \frac{s}{\sqrt{n}} \quad (13)$$

where  $t_{\alpha,v}$  is a value from the Student-t distribution based on significance,  $\alpha$ , and the number of samples,  $n$ . Notice as the number of samples increases, the standard error will decrease and the precision interval collapses onto the true mean.

The density function of the Student-t distribution is given by the following formula:

$$f(t) = \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{k}{2})[\pi v]^{1/2}[\frac{t^2}{v} + 1]^{(v+1)/2}} \quad (14)$$

where  $\Gamma(x)$  is the generalized factorial or the Gamma function.

Reconsider the example of the four drag factor tests. The sample mean was 0.744 with 4 degrees of freedom. The sample standard deviation was 0.0181. Therefore, the standard error is calculated as

$$serr = \frac{s}{\sqrt{n}} = \frac{0.0181}{\sqrt{4}} = 0.00907.$$

If a 95% precision interval is desired, then the mean could change according to Eq. (14). The  $t$  value can be determined using a computer or a table, such as the one found in Refs. (Ayyub and McCuen, 1997; Sheskin, 2000; Montgomery, 2004), to be  $t = 2.776$ . Using this information, the 95% precision interval for the mean of the population of specific drag factors is

$$\mu = 0.744 \pm 2.776(0.00907)$$

which gives a lower bound for the mean of 0.719 and an upper bound for the mean as 0.769. Keep in mind these are the bounds on the mean of the overall distribution of observed drag factors. The actual drag factor a crash event may be outside this range.

### 3.2 Precision Interval of the Sample Variance

Following the argument for the sample mean, the sample variance can also change with subsequent observations of a sample. Assuming the underlying population is normally distributed, the  $\chi^2$  (chi squared) statistic is used to define the precision interval for the population variance. Since the  $\chi^2$  distribution (shown in Fig. 3) is not symmetric, the precision interval must be written as an inequality:

$$P\left(0 < \sigma^2 \leq \frac{(n-1)s^2}{\chi_{\alpha}^2}\right) = 1 - \alpha \quad (15)$$

which says that the probability of the actual population variance,  $\sigma^2$ , falling below  $\frac{(n-1)s^2}{\chi_{\alpha}^2}$  is equal to 1 less the significance level. Again, the values of  $\chi_{\alpha}^2$  can be looked up in a table or computed and they depend on the sample size. Since variance is non-negative, the lower bound for the variance is 0. This is represented by the asymptotic nature of the  $\chi^2$  distribution shown in Fig. 3.

To determine a possible range, only the value from the left side of the  $\chi^2$  distribution is used because this smaller value of  $\chi^2$  gives the maximum limits on the variation. For example, using  $\alpha = 0.05$  and  $n = 4$ , the value of  $\chi_{0.05}^2$  is 0.352. The sample variance from the four drag factor tests is  $3.29 \times 10^{-4}$ . Combining these values in Eq. (15) gives the upper bound of the variance as

$$\sigma^2 \leq \frac{(4-1)(3.29 \times 10^{-4})}{0.352} = 0.00280$$

The number of degrees of freedom has been reduced by 1 for the variance calculations to achieve the favorable statistical quality of being unbiased.

## 4 Applied Statistical Inference

Text books describe the statistical inference procedure when the variance is known or the mean is known (Ayyub and McCuen, 1997). However, when measuring a quantity, both estimators are typically unknown. In the previous section, the precision intervals for both the mean and variance were shown. This section combines both the Student-t and  $\chi^2$  distributions to generate an overall

precision interval for a measured quantity.

## 4.1 The Most Likely Estimate

The sample mean and sample standard deviation are the most likely values for the true population mean and standard deviation. Using only the most likely estimate (rather than a range of probable estimates), a precision interval can be constructed based on the normal distribution.

$$x = \bar{x} \pm z_{\alpha} s \quad (16)$$

where  $\alpha$  is the significance level. For the drag factor example from Section 1.1 the values of  $f$  are computed as

$$f = 0.744 \pm 1.96(0.0181) = 0.744 \pm 0.0355 \quad (95\%)$$

This approach, while recognizing the inherent variability of the underlying measurand,  $x$ , fails to account for the variation of the estimates for the mean and standard deviation.

## 4.2 Variation of the Mean

Accounting for some error in the mean using the Student- $t$  distribution and the standard error gives a larger range than previously reported. When accounting for variation of the mean and using the most likely estimate for the standard deviation, use the following formula to determine a range:

$$x = \bar{x} \pm \left[ (t_{v,\alpha}) \frac{s}{\sqrt{n}} + z_{\alpha} s \right] \quad (17)$$

The example drag factors have 3 degrees of freedom so at the 95% significance level,  $t_{3,95} = 3.182$ . As before,  $z_{95} = 1.96$ . Therefore, the range of drag factor data are:

$$f = 0.744 \pm [3.182(0.0181)/\sqrt{4} + 1.96(0.0181)] = 0.744 \pm 0.0643$$

Notice this interval is expanded from only using the normal distribution and the most likely estimates.

### 4.3 Variation of the Standard Deviation

The  $\chi^2$  distribution represents the variation of the variance according to Eq. (15). If the same logic is used as in the last section, then a range can be determined as:

$$x = \bar{x} \pm z_{\alpha} \sqrt{\frac{(n-1)s^2}{\chi_{n-1}^2}} \quad (18)$$

Again, with 3 degrees of freedom and the 95% confidence interval, the value of  $\chi^2 = 0.352$  (one sided). Applied to the example, Eq. (18) gives the following range for the drag factor:

$$f = 0.744 \pm 1.96 \sqrt{\frac{3(0.0181)^2}{0.352}} = 0.744 \pm 0.104$$

This range is no longer 95% due to the compounding effect of having a 95% normal variate and a 95%  $\chi^2$  value. Furthermore, adding in the effect of an uncertain mean would only push the range of a variable farther apart and increase the confidence level.

For design purposes, this is a conservative approach. However, when performing a reconstruction, or solving an inverse problem, the wide intervals and overestimated precision intervals lead to meaningless results. Therefore, a consistent confidence interval should be maintained while still accounting for all the sources of variability.

### 4.4 Overall Equation

The combination of the mean, the distribution of the mean, and the distribution of the variance gives the following overall equation for the random variable  $X$ :

$$X = \bar{x} + T_{n-1} \left( \frac{s}{\sqrt{n}} \right) + \Phi \sqrt{\frac{(n-1)s^2}{\chi_{n-1}^2}} \quad (19)$$

where  $\bar{x}$ ,  $s$ , and  $s^2$  are statistics from the observed data. The random variables  $T_{n-1}$ ,  $\Phi$ , and  $\chi_{n-1}^2$  are independent. The symbol  $\Phi$  refers to the standard normal variate. The capital or bold letters indicate the variable is random and needs to be represented by a distribution.

Since there is no closed form solution for the distribution of  $X$ , a Monte Carlo simulation is used to determine  $X$ . The Monte Carlo method is well established and has been applied in traffic crash reconstruction (Kost and Werner, 1994; Wood and O’Riordain, 1994; Bartlett, 2003). The basic idea is to independently sample from each random variable and combine the results according to Eq. (19). There must be a sufficient number of samples taken to ensure convergence. The multiple results from Eq. (19) are used to generate an empirical distribution from which a precision interval can be determined.

For example, if 10,000 random samples are taken from each distribution, then constructing the 95% precision interval would require sorting the results and extracting the 250th and 9750th sorted result. These results give the desired range. For the case of the four drag factor tests, the results of the Monte Carlo simulation are shown in Fig. 4. The 95% precision interval shown in Fig. 4 gives a range of the drag factor from 0.67 to 0.83.

## 4.5 Convergence

There exists a most likely distribution that describes the population in question. This is the normal distribution with a mean corresponding to the sample mean,  $\bar{x}$  and a standard deviation corresponding to the sample standard deviation,  $s$ . If only using the most likely estimate, the precision interval would be

$$\underbrace{\bar{x} - (z_{\alpha/2})s}_{MLE\ lower} \leq x \leq \underbrace{\bar{x} + (z_{\alpha/2})s}_{MLE\ upper} \quad (20)$$

where  $z_{\alpha/2}$  is the  $z$ -score from a standard normal distribution corresponding to a precision interval of  $1 - \alpha$ .

The frequentist definition of probability states that probability of an event is the total number of events observed divided by the total of all events. When measuring something, there is no

mathematical bound to the number of times something can be measured. Therefore, the distribution defined by Eq. (19) will converge to the most likely normal distribution as the number of samples increase.

While more mathematically robust proofs for convergence exist, an example will demonstrate the concept of convergence for a realistic example. For this example, consider a 30 random samples from a normal distribution with a mean of 5 and a standard deviation of 0.5. Since these values, shown in Appendix A, come from a known distribution, a sense of the quality of the estimates for the mean and variance can be made.

The graph in Fig. 5 shows the uncertainty associated with lack of knowledge. The interval between the limits of the 95% precision interval of the population and the curved braces is the uncertainty associated with lack of information (i.e. epistemic uncertainty). In a similar fashion, the 95% precision interval gives a representation of the inherent variability of the system (the aleatory uncertainty).

It is important to note that the precision intervals estimated using Eq. (19) always over estimate the interval based on only the estimates from the sample, namely  $\bar{x}$  and  $s$ . As the number of samples increase, the difference between the results of Eq. 19 and Eq. (20) will reduce as seen in Fig. 5.

## 4.6 Discussion of Results

A Monte Carlo simulation is capable of rendering very accurate *mathematical* results given sufficient sampling. However, the simulation does not police its applicability to the assumptions made when constructing the model. The measurements taken must accurately represent the physics of the problem.

An important assumption of this analysis is that of normality. With a truly Gaussian (normal) variable, the possibility exists for a value to take on an unrealistic value. This is especially true for quantities that have a large amount of variation. For example, analyzing stiffness coefficients from crash test data may produce a range of negative stiffness. In such unrealistic situations, the source data must be reexamined and the results of a statistical analysis must be tempered by the laws of

physics.

The results of the numerical example in the previous section suggest that three samples may not be sufficient to eliminate a significant amount of epistemic uncertainty. The epistemic uncertainty is reduced by half by taking one more sample. Steady improvements are made in reducing the epistemic uncertainty as the number of samples increase from 5 to 10. The convergence of the estimated precision intervals to the actual precision interval is slower after 10 samples. Therefore, it is highly recommended to use four or more samples when determining a precision interval (range) for a measured quantity.

## **5 Conclusions**

This paper provided motivation for using a range for use in traffic crash reconstruction analysis. There is a need to have a mathematically consistent range (precision interval) for parameters that are estimated with a small number of measurements. This precision (typically 95%) should be maintained while accounting for all sources of uncertainty. The precision interval was calculated using sampling statistics and the Monte Carlo integration scheme. An example of a convergence study suggests that at least four measurements should be made to estimate an unknown quantity to reduce the epistemic uncertainty to tolerable levels.

## **Acknowledgments**

Appreciation should be extended to Dr. Brian Beachkofski and Mr. Jeff Brown at the U.S. Air Force Research Laboratory for their insightful discussions regarding this topic. John Daily, Andy Rich, and Bill Messerschmidt provided insightful suggestions. Also, acknowledge the Illinois Association of Technical Accident Investigators (IATAI) for requesting this topic be explored.

## **Notation**

The following symbols are used in this paper:

Symbol	Meaning
$\bar{x}$	= sample mean
$\mu$	= population mean
$s$	= sample standard deviation
$\sigma$	= population standard deviation
$f(\cdot)$	= Functional form for a probability density function
$f$	= drag factor
$F(\cdot)$	= functional form for a cumulative distribution function
$x$	= variable, measurand
$X$	= random variable
$n$	= number of samples
$\nu$	= degrees of freedom
$serr$	= standard deviation of the mean
$\alpha$	= significance level
$z$	= z-score for a standard normal distribution
$T_\nu$	= student- $t$ distribution with $\nu$ degrees of freedom
$\Phi$	= the standard normal variate
$\chi^2$	= the chi-squared distribution

## References

- Aitken, C. G. G. and Taroni, F. (2004). *Statistics and the Evaluation of Evidence for Forensic Scientists*. John Wiley & Sons, Chichester.
- Ayyub, B. M. and McCuen, R. H. (1997). *Probability, Statistics, and Reliability for Engineers*. CRC Press LLC, Boca Raton.
- Bartlett, W. (2003). Conducting monte Carlo analysis with spreadsheet programs. *SAE Technical Paper Series*, (2003-01-0487).

- Bartlett, W. et al. (2002). Evaluating the uncertainty in various measurement tasks common to accident reconstruction. In *Accident Reconstruction SP-1666*, number 2002-01-0546 in SP, pages 57–70. Society of Automotive Engineers, Warrendale, PA.
- Daily, J. G. et al. (2006). *Fundamentals of Traffic Crash Reconstruction*. Institute of Police Technology and Management, University of North Florida, Jacksonville, FL.
- Dieck, R. H. (2002). *Measurement Uncertainty: Methods and Applications*. The Instrumentation, Systems and Automation Society, Research Triangle Park, NC, third edition.
- Efron, B. (1982). *The Jackknife, the Bootstrap and other Resampling Plans*. Society for Industrial and Applied Mathematics, Philadelphia.
- Kost, G. and Werner, S. M. (1994). Use of Monte Carlo simulation techniques in accident reconstruction. *SAE Technical Paper Series*, (940719).
- Kreyszig, E. (1970). *Introductory Mathematical Statistics*. John Wiley, New York.
- Marks, C. C. O. (2002). Accident analysis uncertainty in the forensic context. Personal Communication.
- Montgomery, D. C. (2004). *Design and Analysis of Experiments*. John Wiley and Sons, New York, 5th edition. A complete discussion of designing, conducting, and analyzing experiments.
- Oberkampf, W. L. et al. (2001). Mathematical representation of uncertainty. *American Institute of Aeronautics and Astronautics (AIAA)*, (1645):1–22.
- Rivers, R. W. (1997). *Advanced Traffic Accident Investigation*. Institute of Police Technology and Management, Jacksonville.
- Roberson, B. and Vignaux, G. A. (1995). *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*. John Wiley & Sons, Chichester.

Sheskin, D. J. (2000). *Handbook of Parapmetric and Nonparametric Statistical Procedures*. Chapman & Hall / CRC, Boca Raton.

Weisstein, E. W. (2006). Monte carlo method. From MathWorld—A Wolfram Web Resource. <http://mathworld.wolfram.com/MonteCarloMethod.html>.

Wood, D. P. and O’Riordain, S. (1994). Monte carlo simulation methods applied to accident reconstruction and avoidance analysis. *SAE Technical Paper Series*, (940720).

## **A Random Samples**

The following is a list of random samples from a normal distribution with a mean of 5 and a standard deviation of 0.5. These values were used in the analysis to generate Fig. 5. The values used are 5.062, 5.523, 4.607, 5.312, 5.345, 4.237, 4.819, 4.735, 4.815, 5.622, 5.187, 4.940, 4.800, 4.911, 5.675, 4.497, 4.788, 3.770, 4.866, 5.342, 4.849, 4.292, 5.542, 5.127, 4.006, 4.318, 5.779, 5.070, 5.003, and 5.580. The 95% precision interval for the underlying population is

$$4.020 < x < 5.980$$

as shown by the dotted lines of Fig. 5.

## List of Figures

1	The probability functions for normal distribution with a mean of $\mu$ . The gray area under the curve represents 95% of the total area. . . . .	21
2	Examples of the Student- $t$ distribution. The dashed line has 2 degrees of freedom, the solid line has 5 d.o.f. and the dotted line is the standard normal. . . . .	22
3	Examples of the $\chi^2$ distribution. . . . .	23
4	The result of the Monte Carlo simulation is shown as an empirical CDF for the four drag factor tests considered in the example. The actual value of the drag factor has a 95% chance of lying in the range shown above. . . . .	24
5	Convergence of Eq. (19) is demonstrated by the precision intervals drawn with the curved ends. Convergence of the range predicted by the normal distribution based on the sample mean and standard deviation is shown with the square ends. The convergence of the mean is demonstrated with the dots. . . . .	25

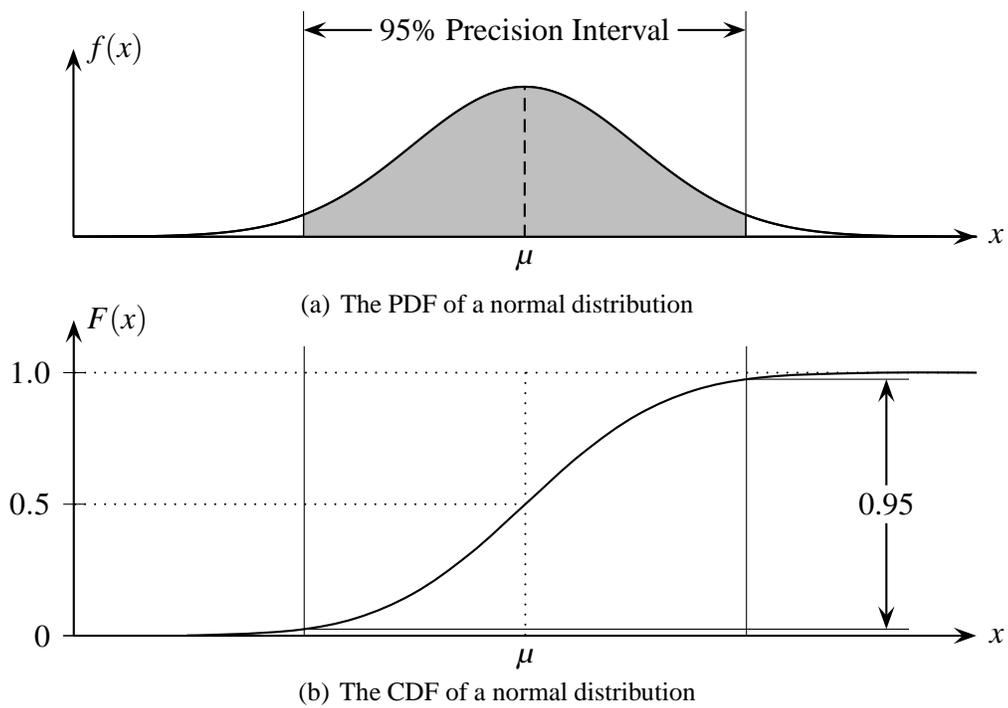


Figure 1: The probability functions for normal distribution with a mean of  $\mu$ . The gray area under the curve represents 95% of the total area.

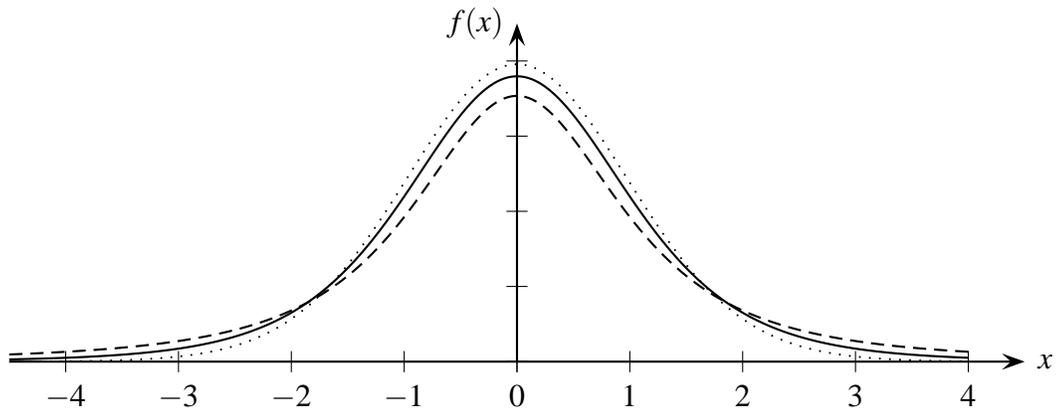


Figure 2: Examples of the Student- $t$  distribution. The dashed line has 2 degrees of freedom, the solid line has 5 d.o.f. and the dotted line is the standard normal.

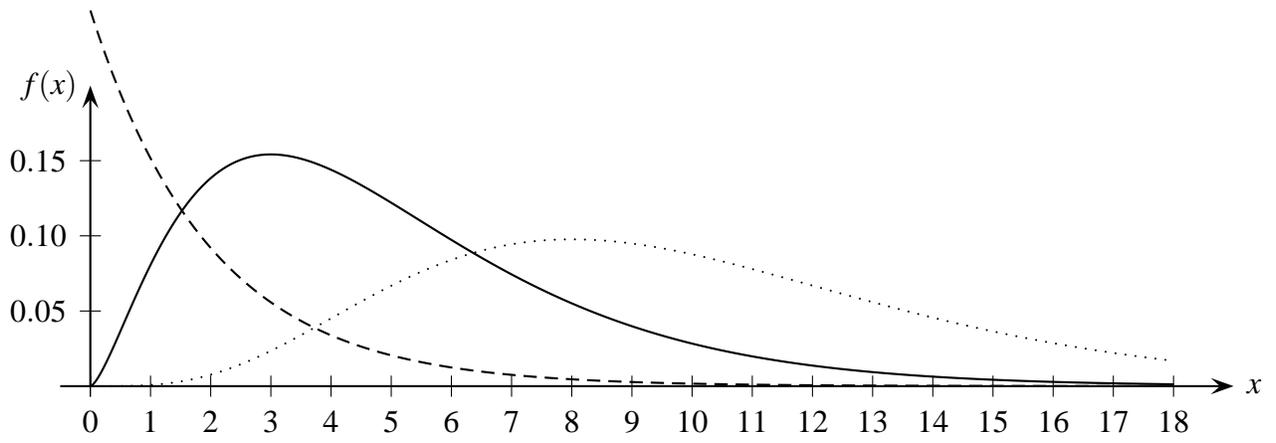


Figure 3: Examples of the  $\chi^2$  distribution.

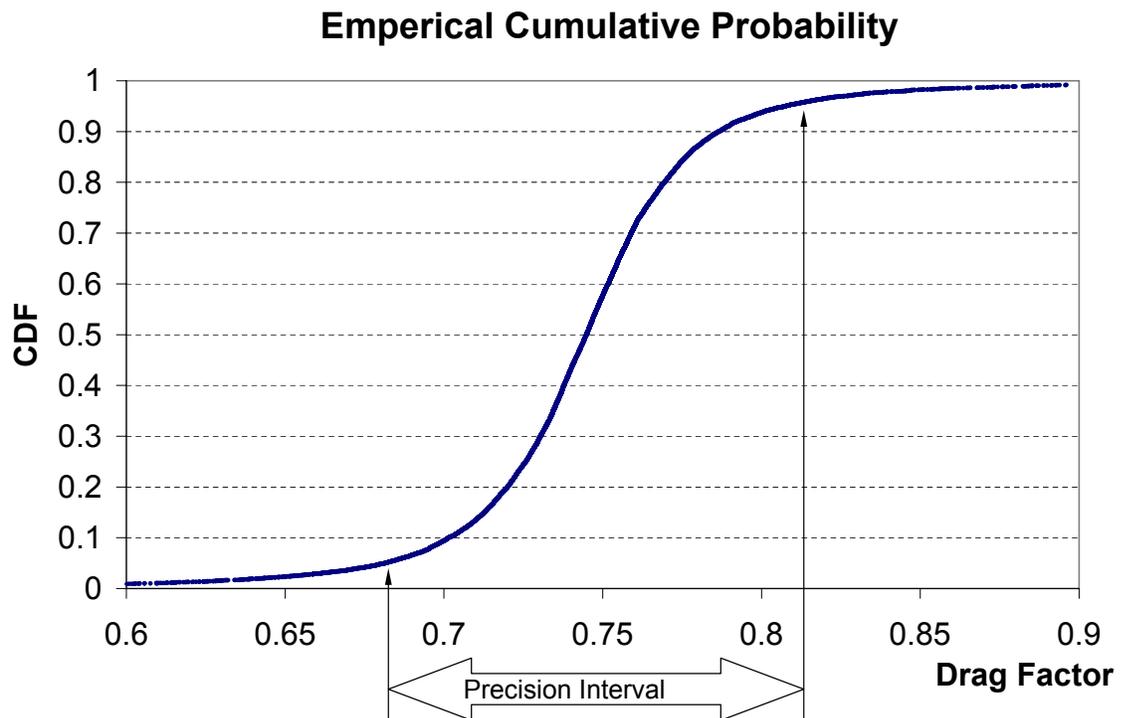


Figure 4: The result of the Monte Carlo simulation is shown as an empirical CDF for the four drag factor tests considered in the example. The actual value of the drag factor has a 95% chance of lying in the range shown above.

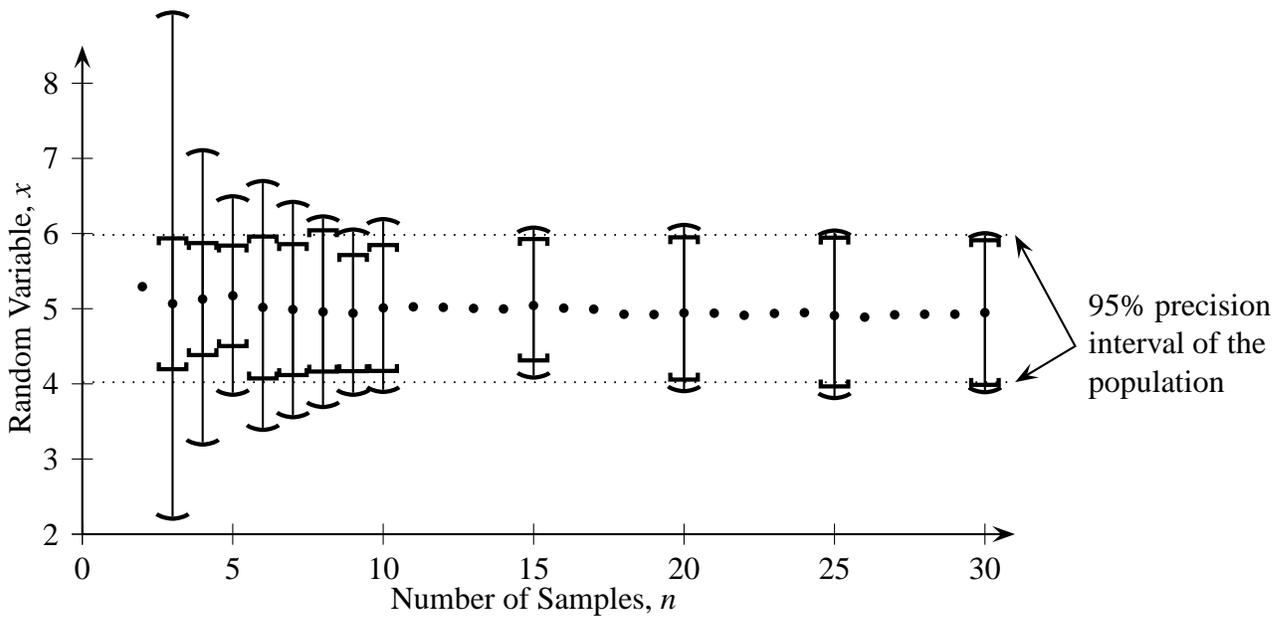


Figure 5: Convergence of Eq. (19) is demonstrated by the precision intervals drawn with the curved ends. Convergence of the range predicted by the normal distribution based on the sample mean and standard deviation is shown with the square ends. The convergence of the mean is demonstrated with the dots.