Determining confidence intervals, and convergence, for parameters in stochastic evacuation models

ANGUS GRANDISON*
University of Greenwich
Old Royal Naval College
Park Row
Greenwich
LONDON SE10 9LS, UK
*Corresponding Author: A.J.Grandison@gre.ac.uk 020 8331 7912

https://orcid.org/0000-0002-9714-1605

Abstract

An issue when using stochastic egress models is *how many simulations are required to accurately represent the modelled scenario?* Engineers are mostly interested in a representative Total Evacuation Time (TET). However, the convergence of the TET may not ensure that the full range of evacuation dynamics has been adequately represented. The average total egress curve (**AC**) has been suggested as an improved measure. Unfortunately, defining a confidence interval (CI) for the **AC** is problematic. CIs can robustly quantify the precision of many statistics and have been used to define convergence in egress modelling and other research fields. This paper presents a novel application of bootstrapping, functional analysis measures (FAMs), and a bisection algorithm, to derive three FAM-based CIs representing the precision of the **AC**. These CIs were tested using a theoretical model to demonstrate the consistency of the coverage probability, the actual percentage of CIs that contain the theoretical parameter, with the nominal 95% confidence level (NCL). For two of the FAM-based CIs, it was found that the coverage probability was between 94.2% and 95.6% for all tested sample sizes between 10 to 4000 simulations. The third FAM-based CI's coverage probability was always greater than the NCL and was a conservative estimate, but this presented no problems in practice. A FAM-based CI may suggest if there is more or less variability in an earlier phase of the evacuation. A convergence scheme based on statistical precision, CI widths, is proposed and verified. The method can be extended to other statistics.

Acronyms

CI	confidence interval
CL	confidence level
TET	total evacuation time
MT	mean TET
SD	standard deviation (of TETs)
CDF	Cumulative Density Function
BCa	Bias Corrected and accelerated (bootstrap)
FAM	functional analysis measure
ERD	Euclidean Relative Distance
EPC	Euclidean Projection Coefficient
SC	Secant Cosine

Symbols

a, **b**	egress curves (an ordered vector of agent evacuation times)
AC	average egress curve
$(1-\alpha)$	confidence level value (expressed as a decimal)
B	number of bootstrap resamples
p	percentile (expressed as a decimal)
$\acute{p}$	"Hesterberg" corrected percentile
m	number of agents/people in a simulation/experiment

n	sample size (generally the number of simulations or experiments)
df	degrees of freedom (n-1)
$T^{-1}(p, df)$	the inverse CDF of Student's t-distribution for p
$P(x)$	the probability that x occurs.
$\Phi(z)$	the CDF of the normal distribution up to z
$\Phi^{-1}(p)$	the inverse CDF of the normal distribution for p
$\lfloor x \rfloor$	integer floor of x
$\lceil x \rceil$	integer ceiling of x
$\lvert x \rvert$	absolute value of x
θ	statistic of interest. e.g. MT, SD, ERD, EPC, and SC
$\theta^{(n)}$	statistic of interest calculated from n values
$\theta^{(\infty)}$	population parameter of interest
θ_{-i}	θ calculated using the original n simulations barring the ith simulation.
θ^*	bootstrap statistic of interest
θ_j^*	j^{th} bootstrap statistic of the (unsorted) distribution of θ^*
$\theta_b^{*,o}$	b^{th} bootstrap statistic of the (ascending) ordered distribution of θ^*
W_θ	normalised width of CI_θ
$\forall k$	for all agents k = 1 to m

Subscripts

l	lower CI limit
u	upper CI limit

1 Introduction

A stochastic approach to evacuation simulation [1, 2] is employed in many models [3-8] to reflect uncertainty in human behaviour [9]. If an (experimental or simulated) evacuation trial is repeated using the same population and the same initial conditions, it is possible that the evacuation will progress differently and may result in differences in total evacuation time (TET) or other statistics of interest. In addition to that uncertainty there is also uncertainty related to the initial conditions where the distribution of agents (e.g. number of agents and starting locations) and their attributes (e.g. walking speeds and response times) may also cause differences between trials. These uncertainties can be simulated by randomly sampling suitable distribution functions for those attributes. A simple random approach is adopted in Monte Carlo egress simulation although other possible sampling methods could be utilised [10].

Many simulations are needed to represent the range of potential behaviours which may be expected for a particular scenario. However, this is set against the potentially significant computational cost of running a stochastic egress model which an end user would wish to minimise. Thus, a major issue facing stochastic egress model users is determining when enough simulations have been performed to give an accurate representation of the modelled scenario and the statistics of interest. In evacuation modelling literature [11-20] that examines this issue there are broadly two approaches suggested. The first approach is based on running a fixed number of simulations. The second approach is based on some dynamic assessment of the behaviour of the statistic(s) of interest.

When the first approach is used, a wide range for the required number of simulations has been suggested. For high speed train evacuation analysis, Li et al. [11] suggested that 10 simulations would suffice provided certain conditions were met. For large passenger ship analysis, the International Maritime Organisation (IMO) recommends that a minimum of 500 simulations should be used to determine the 95th percentile TET [12]. They additionally stated that a lower number of simulations would be allowable provided a suitable convergence methodology was followed [13]. For aircraft evacuation analysis, Galea [14] noted that 1000 simulations could typically be utilised to determine the 95th percentile TET. Meacham et al. [15] utilised 2000 simulations to construct a Cumulative Density

Function (CDF) of the TETs for building scenarios. There is no consensus on the number of simulations used and there is generally little justification for the number of simulations proposed. The disadvantage of this approach is that it is unknown whether the number of simulations is sufficient to represent the range of behaviours that may exist. Furthermore, the number of simulations performed could be excessive, for the intended purpose, which could limit the number of potential scenarios or alternative designs that could be explored. This approach does have the advantage of requiring little additional modeller expertise but is also limited as some models may be very sensitive to the chosen number of simulations [16].

In the second approach, additional egress simulations are performed until some indicators of convergence have been satisfied. With this approach, the number of simulations performed is unknown a priori. The eventual number of simulations performed is based on the behaviour of the statistics of interest and should ensure that sufficient, and not excessive, computation being performed. Compared to running an arbitrary number of simulations this requires some expertise in determining suitable convergence tolerances for the scenario being run. Two types of indicators have been proposed, the first type is based on confidence intervals (CIs) [21, 22] and the second type is based on examining successive differences in the statistics of interest with additional simulation.

CI based techniques have been successively used in egress modelling [13, 20], and various other fields of study [23-28], to define convergence. For large passenger ship analysis, Grandison et al. [13] calculate the CI of the 95th percentile TET, using a binomial distribution, and incrementally test, with additional simulations, the limits of the CI against a pass/fail criterion [12]. This technique was found to accurately determine whether a ship design had passed or failed, whilst minimising the computational effort required. Furthermore, they noted that this technique could be adapted to determine convergence based on a required precision and that the methodology could be extended to other suitable statistics of interest and types of egress analysis. A disadvantage of this method is that it can be difficult to determine CIs for some statistics. However, an advantage of this scheme is that the convergence indicators are directly linked to the precision of the statistics of interest.

A successive difference technique has been proposed by Ronchi et al. [17] and Ronchi & Nilsson [18]. Ronchi et al. [17] define convergence (i.e. sufficient repeat simulations have been performed) when their five statistics of interest (MT, SD and three functional analysis measures (FAMs) of the average total egress curve (**AC**)) were all changing by less than their specified tolerances per simulation over a set number of simulations. It was suggested that these tolerances could be based on the uncertainty of the available safe egress time from a fire modelling analysis. Additionally, Lovreglio et al. [19] proposed that those tolerances could be based on values determined from a set of repeated evacuation experiments when used to validate a stochastic egress model. An advantage of the successive difference approach is the simplicity in calculating the convergence indicators. However, Ronchi et al. noted that '*The first limitation of the method is that it uses the concepts of convergence in mean and the central limit theorem rather than a statistical estimation of the expected values...*'. This means that their convergence indicators do not directly relate to standard errors or CIs of the statistics of interest.

From the previous examples, the parameter of most interest is a representative TET. However, an analysis based solely on the TET may miss important details of the evacuation [16-20, 29], e.g., a set of egress trials could have similar representative TETs, but the individual evacuation dynamics could be very different. A more detailed analysis of the evacuation can be performed by additionally examining the exit time of certain percentages of the agents [18, 20]. However, an even greater level of detail can be achieved by examining the entire agent egress curve (an ordered vector of the individual agent's exit time) via FAMs [16, 17, 19, 29]. Furthermore, egress curves are a typical output from many evacuation models. The convergence behaviour of the **AC** was used to represent behavioural uncertainty, the uncertainty associated with the stochastic nature of human behaviour, by Ronchi et al.

CIs have been demonstrated to accurately and efficiently determine convergence for statistics of interest in egress modelling and other fields of study. Ronchi et al.'s use of the **AC** was an important innovation in studying egress variability. It is therefore desirable to extend their original method by applying CIs

to the **AC**, thereby giving a more standard statistical interpretation of behavioural uncertainty. However, it is not intuitively obvious how to determine the CIs for this curve and no solution has been previously published. The key contribution of this paper is the development of FAM-based CIs for the **AC** (section 3) with an accurate overall confidence level. This was achieved using the novel application of bootstrapping (see section 2.1), FAMs (see section 2.2), and a bisection algorithm (see section 3.1). This addresses the first limitation of Ronchi et al.'s method and allows the benefits of the CI approach to convergence to be combined with the **AC** within a common framework. It is also advantageous to use the FAMs, to form the CIs, as they are somewhat familiar to the fire and evacuation research communities. FAMs were introduced to the fire safety field in 1999 [30] and the first reported use for evacuation was in 2012 [29]. A further contribution of this paper is the development of a convergence scheme. The convergence scheme is based on the same variables used by Ronchi et al. [17], i.e. **AC**, MT, and SD, but utilises CIs. The CI for the MT is based on the standard t-distribution expression (see section 2) and the CI for the SD is obtained using a standard application of bootstrapping (see section 2.1.1). This convergence scheme will efficiently determine when all the statistics of interest have reached their required statistical precisions. Utilising convergence of the **AC** helps to ensure that variability, between simulations, which is not expressed by examining TETs alone, is also captured.

The CIs and the convergence scheme are tested and verified using a simple theoretical model. Apart from demonstrating the correctness of the methods, that work also feeds into a discussion on how to determine appropriate tolerances for the FAM-based CIs of **AC**.

2 Statistical and mathematical background

This section provides a brief background on the necessary terminology and methods used in this paper. Initially, some standard statistical terminology will be described; this initial section can be skipped over by readers with a grounding in statistics. This is followed by an overview of "bootstrapping" with some worked examples (section 2.1). Finally, a summary of functional analysis, used to compare egress curves, is given (section 2.2).

A sample *statistic* is the value of an *estimator* based on a sample from the population of results. A population *parameter* can be considered as the value of an *estimator* based on the entire population of results. Examples of estimators include the mean TET (MT) (eq. 1), the 95th percentile TET, and the standard deviation (SD) of TET. Ideally, one would want to know the parameter but practically it is (generally) only possible to obtain the statistic of a sample of results and a range of potential values for the parameter. The sample statistic would display variability depending on the nature of the sample obtained although the sample statistic tends to the value of the parameter as the size of the sample increases (Law of Large Numbers [22]).

$$MT = \frac{\sum_{i=1}^{n} TET_i}{n} \tag{1}$$

The "expected value" of a sample statistic is the mean value of a series of infinitely repeated independent instances of that sample statistic with the same sample size [22]. This will equate to the corresponding population parameter if the estimator is unbiased. This is the case for many estimators although the sample SD is slightly biased. Whilst the standard variance bias correction has been applied (eq. 2), i.e. dividing by $(n-1)$ rather than n, there is still some residual bias. It is reasonable to neglect this bias given the large scale of the variability associated with the SD compared to the small size of the bias and the complexity of calculating that bias. The small bias that does exist also diminishes as the sample size increases.

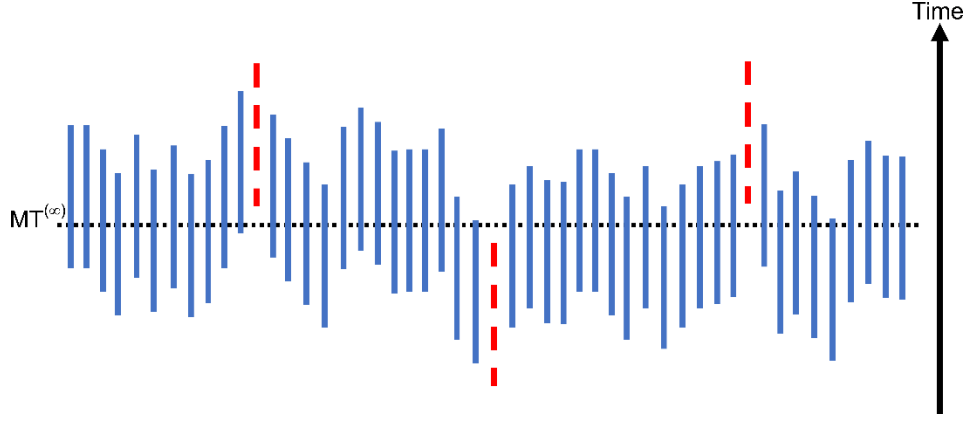$$SD = \sqrt{\frac{\sum_{i=1}^{n} (TET_i - MT)^2}{(n-1)}} \tag{2}$$

4

Fig. 1 - Example 95% Cls for population parameter $MT^{(\infty)}$ for 50 different samples. Forty-seven (solid) CIs contain the population parameter and three (dashed) CIs do not contain the population parameter.

Although it is generally impossible to exactly know the population parameter it is possible to estimate a range of possible values for the population parameter based on the sample of results. This is a CI, which will differ between different samples but will frequently include the parameter of interest. Ideally, this frequency is determined by the confidence level (CL). However, the actual frequency, known as the coverage probability, may differ in practice. A set of CIs for 50 different samples is depicted in Fig. 1. This is illustrated for a 95% CL in Fig. 2 where the vast majority, 47 (~95%), of the sample CIs contain $MT^{(\infty)}$ but 3 (~5%) of the sample CIs do not. The superscript brackets indicate the number of samples used to calculate the statistic, when the number is ∞ the statistic is equivalent to the population parameter. In general, only a single sample of results will be generated, although that single sample could contain hundreds of simulations, so only one CI will be generated per statistic of interest.

The calculation of the CI for certain parameters is straightforward due to the availability of standard expressions. For the MT, the CI with a 100(1-α)% CL can be calculated using eq. 3 where $T^{-1}(p, df)$ is the inverse t-distribution for percentile p with df degrees of freedom [22]. This expression assumes that the population SD of the TETs is unknown, which is generally the case in egress models/experiments, and is estimated with the sample SD. If the original data is normally distributed, then the expression is valid for $n > 1$. For non-normally distributed data, the expression is generally valid due to the central limit theorem for a larger n but is dependent on the nature of the original data distribution. Some sources advocate a minimum n of 30 although for skewed data a minimum n of 40 [31] has been suggested. It is further assumed that the data is independent and identically distributed.

$$100(1-\alpha)\% \; CI = \left[\left(MT^{(n)} + T^{-1}\left(\frac{\alpha}{2}, n-1\right) \times \frac{SD^{(n)}}{\sqrt{n}} \right), \left(MT^{(n)} + T^{-1}\left(1-\frac{\alpha}{2}, n-1\right) \times \frac{SD^{(n)}}{\sqrt{n}} \right) \right] \qquad (3)$$

In eq. 3, the CI is quoted as a lower limit and an upper limit of the population parameter value, e.g. 150.0 with a 95% CI [135.0, 165.0]. The CI could also be quoted as the difference between the limits and the sample statistic, e.g. 150.0 with a 95% CI [-15.0, +15.0]. Those differences could also be quoted as percentages, e.g. 150.0 with a 95% CI [-10%, +10%]. If the CI is symmetric then it can be expressed in shortened form with a plus-minus limit, e.g. 150 ± 10% (95% CI).

A CI is a trade-off between its width, which should ideally be as narrow as possible, and its CL, which should ideally be as high as possible. Ideally, one would want a 100% confidence, but this leads to an infinitely wide CI which is not very useful. For a statistic calculated from a fixed number of data points, the CI width will increase with increasing CL. Having a very high CL will result in a very wide CI whilst a "low" CL with a correspondingly narrow CI is also unhelpful as there is little confidence that the CI contains the parameter of interest. A 95% CL is typically chosen as it is very likely to contain the population parameter whilst maintaining a relatively narrow width. When calculating the CI_{MT} (eq. 3) with varying CLs (see Table 1), a 99% CL gains an extra 4% in confidence but requires a 33%

increase in CI width compared to a 95% CL. A higher 99.9% CL gains an extra 4.9% in confidence but requires a 71% increase in CI width compared to a 95% CL. Using a lower 68% CL (equivalent to 2 × standard error widths) reduces the CI width by 49% but there is a reduction in confidence of 27% compared to a 95% CL. From eq. 3, the width (W_{MT}) of the CI_{MT} with a fixed CL will also tend to narrow with increasing n (see eq. 4).

$$W_{MT}(n) \propto \frac{1}{\sqrt{n}} \tag{4}$$

Table 1 - Calculation of the relative widths of CIs for the mean value (eq. 3) with increasing CL for a sample size of 100.

CL	68%	95%	99%	99.9%	100%
CI width relative to 2 × standard error	1	1.98	2.63	3.39	∞

In order to further discuss CIs, it is necessary to describe *sampling* distributions. A *sampling* distribution is the probability distribution of a statistic for a fixed sample size for all the possible samples that could be selected [22]. For example, a set of five TETs are obtained from repeated trials and the statistic of interest is the mean of those TETs. If this set of five trials is repeated, a further set of TETs can be obtained, and another mean of those values is calculated. Each time the process is repeated a different estimate for the mean value will be obtained, see Table 2.

Table 2 - An example set of samples taken from a distribution. Each sample consists of five values.

Sample	Sample TETs	Mean TET
1	133, 127, 148, 140, 177	145
2	161, 124, 159, 145, 155	148.8
3	173, 147, 167, 134, 178	159.8
…	…	…
$n \rightarrow \infty$	129, 139, 172, 113, 116	133.8

If this could be repeated a very large number of times, then it would be possible to produce a distribution of the sample mean values, i.e. the sampling distribution. Assuming the TETs are normally distributed with an unknown population SD then it has been shown theoretically that the sampling distribution for the example can be represented by a t-distribution [22]. The mean of the sampling distribution is equal to the expected value and the population mean of the original TET distribution. From the central limit theorem, the standard deviation of the sampling distribution, also known as the standard error, is the standard deviation of the original TET distribution divided by $\sqrt{n}$.

From this sampling distribution, the interval that contains the innermost $100(1 - \alpha)$% of sample means can be described by eq. 5 (see Fig. 2).
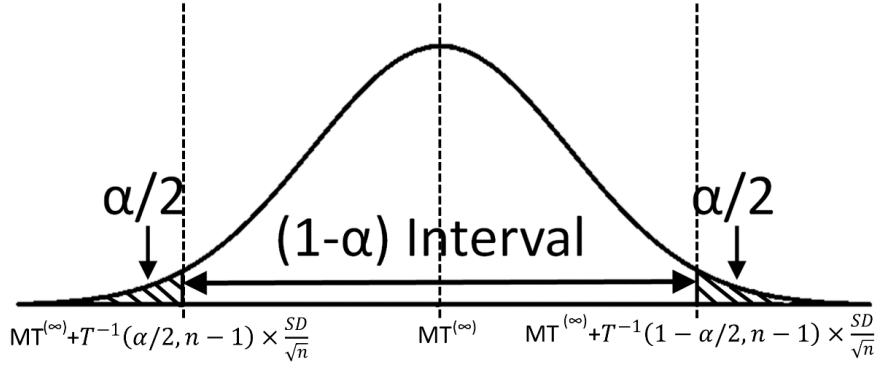
Fig. 2 - Example sampling distribution with a mean of $MT^{(\infty)}$ and standard deviation of $SD/\sqrt{n}$. The interval containing the central $100(1-\alpha)\%$ sample means is depicted.

$$P\left(\left\{MT^{(\infty)}+T^{-1}\left(\frac{\alpha}{2},n-1\right)\times\frac{SD}{\sqrt{n}}\right\} < MT^{(n)} < \left\{MT^{(\infty)}+T^{-1}\left(1-\frac{\alpha}{2},n-1\right)\times\frac{SD}{\sqrt{n}}\right\}\right) = 1-\alpha \quad (5)$$

Equation 5 is not particularly useful as in general the value of the $MT^{(n)}$ is known and the $MT^{(\infty)}$ is unknown. However, eq. 5 can be algebraically recast [22] to make the $MT^{(\infty)}$ the subject of the inequality leading to the probability that the $MT^{(\infty)}$ lies within a CI with a CL of $100(1-\alpha)\%$ (eq. 6). This is a probabilistic statement of the CI given in eq. 3.

$$P\left(\left\{MT^{(n)}+T^{-1}\left(\frac{\alpha}{2},n-1\right)\times\frac{SD}{\sqrt{n}}\right\} < MT^{(\infty)} < \left\{MT^{(n)}+T^{-1}\left(1-\frac{\alpha}{2},n-1\right)\times\frac{SD}{\sqrt{n}}\right\}\right) = 1-\alpha \quad (6)$$

For some statistics, such as the mean value (eq. 3) and the 95th percentile value [13], a theoretical solution exists for determining the CI. However, for many statistics a theoretical solution may not exist or is difficult to derive. For those cases, it may be possible to use bootstrapping to approximate the sampling distribution and hence derive the CI.

2.1 Bootstrapping concepts

Bootstrapping [32-36] is a widely used computational statistical technique. The premise of bootstrapping is that inference about a *population* from sample data can be modelled by resampling the sample data and performing inference about a sample from resampled data. Generally, the population distribution is unknown and only the sample of data is known. Bootstrapping uses this sample of data as an approximation to the true population distribution. The bootstrapping technique is applicable to many estimators although there are known difficulties associated with the modal value [37]. Care is also required when applying the technique to heavy-tailed distributions particularly those with infinite variance [38, 39]; this is not anticipated to be a major issue in the egress modelling context.

A simple worked example, to determine the CI of the mean, follows. In practice, bootstrapping is not generally used to determine the CI of a mean value due to the wide applicability of eq. 3. The primary intent of this example is to demonstrate the methodology. A set of *n* values are sampled from some unknown distribution and leads to the set of values in the "Original" row of Table 3. In this example, *n* is set to five. This data could be expensive to obtain, e.g. these could be TETs from a set of egress trials. The mean of the original sample is 139.5. The original sample is now resampled to create a bootstrap sample, which is a very inexpensive process. Each value in the original sample has the same, $1/n$, chance of being selected. Values are randomly copied from the original sample *n* times to create a bootstrap sample (see Table 3). This implies that a particular value from the original sample may be represented once, more than once, or not at all in a bootstrap sample. Furthermore, only values from the original sample can be represented in the bootstrap sample. For example, it can be seen for bootstrap 1 that 99.7 is represented twice and 188.5 is not represented at all and the other values are represented once each. When *n* values have been selected, the mean value of the bootstrap sample is calculated.

This process is repeated *B* times, where *B* >> *n*, to create a bootstrap *sampling* distribution of the mean (cf. Table 2). In this example *B* is set to 19. This is a far lower number of bootstrap samples than would be used in practice but is used here for demonstration purposes.

Table 3 - A worked example of bootstrap samples based on an original sample of five TETs. Only select bootstraps are listed for compactness.

Sample	Sample TETs	Mean TET
Original	134.0, 188.5, 149.5, 99.7, 126.1	139.5
Bootstrap 1	126.1, 149.5, 99.7, 99.7, 134.0	121.8
Bootstrap 2	188.5, 188.5, 149.5, 99.7, 99.7	145.2
…	…	…
Bootstrap 6	126.1, 188.5, 188.5, 188.5, 134.0	165.1
…	…	…
Bootstrap 10	126.1, 188.5, 188.5, 188.5, 188.5	176.0
…	…	…
Bootstrap 15	126.1, 134.0, 99.7, 126.1, 99.7	117.1
…	…	…
Bootstrap 19	149.5, 149.5, 134.0, 126.1, 126.1	137.0

The CI, with a 90% CL (the maximum CL that can be represented with 19 bootstraps), is obtained by taking the percentile values that bound the innermost 90% of the sampling distribution which are the 5th and 95th percentile values. The bth bootstrap index, of the vector of ascending ordered bootstrap means, representing the $(100p)^{th}$ percentile is given by eq. 7.

$$b = (B + 1)p \qquad (7)$$

This equates to the 1st and 19th values of the ordered bootstrap sampling distribution of the mean (see Table 4). The bootstrap 90% CI of the mean is [117.1, 176.0]. Using the conventional calculation (eq. 3) the 90% CI is [108.3, 170.8]. The bootstrap CI is not particularly close to the standard approach although is not unreasonable given the limited sample size and small number of bootstraps. The accuracy of the method significantly improves as *n* and *B* increase, and as more sophisticated methods are used to select the required percentiles from the bootstrap sampling distribution.

Table 4 - The lower and upper ends of the ordered bootstrap sampling distribution of the mean for the worked example.

Percentile	0	5	10	…	90	95	100
Ordered Index (eq. 6)	n/a	1	2	…	18	19	n/a
Original Index	n/a	15	1	…	6	10	n/a
Mean value	$-\infty$	117.1	121.8	…	165.1	176.0	∞

The above worked example was repeated using a sample size of 100 values and 1999 bootstraps. In this case the mean of the sample was 156.5. The bootstrap 90% CI, [151.1, 161.5], gives a good approximation to the conventionally determined 90% CI (eq. 3), [151.4, 161.6], in this instance. The bootstrap sampling distribution of the mean for this case is presented as a histogram in Fig. 3. The bootstrap mean sampling distribution is similar in shape to a normal distribution and approximates the true sampling distribution.
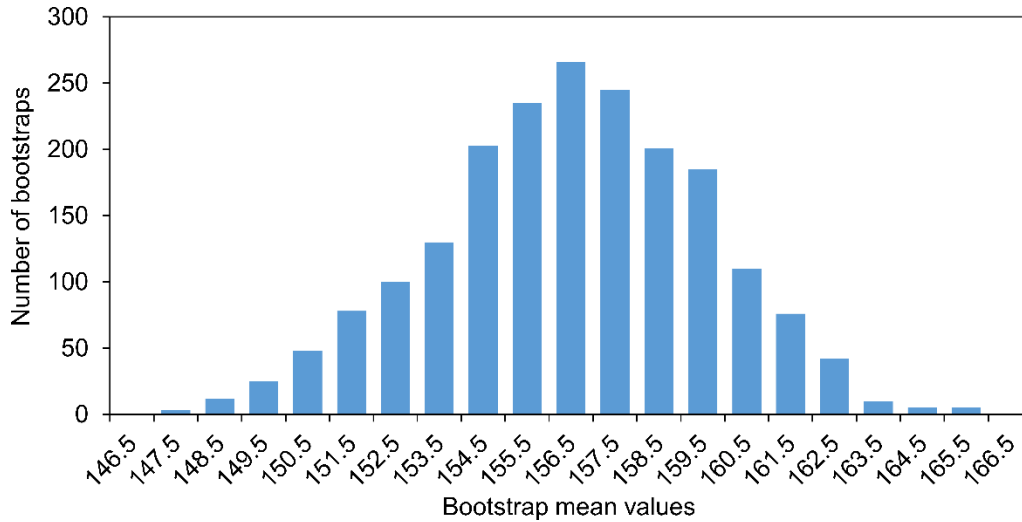
Fig. 3 – Histogram of 1999 bootstrap mean TETs for the worked example. The sample size was 100.

The simplest approach for selecting a centred CI with a $100(1 - \alpha)\%$ confidence level is to assume it is bounded by the $100(\alpha/2)^{th}$ and $100(1 - \alpha/2)^{th}$ percentile bootstrap samples. This is the approach used for the worked example above. However, this simple percentile approach has two notable problems. Firstly, it tends to give narrower CIs than expected when "small" sample sizes were used. Secondly, it does not account for skew and bias that may exist in the samples.

For the first problem, Hesterberg proposed a correction based on the normal distribution and the t-distribution [40]. This is based on correcting the percentile limits of the bootstrap sampling distribution of the mean. The "corrected" $(100\acute{p})^{th}$ percentile based on the original $(100p)^{th}$ percentile of sample size n is given by eq. 8, where $\Phi(z)$ is the CDF of the standard normal distribution function. This correction always tends to make the CI more conservative but could result in an over-correction or insufficient correction of the required percentile dependent on the actual, but unknown, form of the sampling distribution. In the event of an over-correction, the CI will be wider than it should be but represents a more conservative estimate. If the CI is insufficiently corrected, then at least some correction is applied compared to no correction thus bringing the coverage probability of the CI closer to the nominal confidence level.

$$\acute{p}(p,n) = \Phi\left(\sqrt{\frac{n}{n-1}}\,\mathrm{T}^{-1}(p, n-1)\right) \tag{8}$$

Table 5 illustrates the correction (eq. 8) applied to the 5th percentile for various sample sizes. The effect of the correction is significant for small n but diminishes as n increases.

Table 5 - The value of the corrected 5th percentile (eq. 8) based on sample size n.

n	10	20	30	40	100	1000
$100\acute{p}$	2.6	3.8	4.2	4.4	4.8	5.0

For the second problem associated with possible skew and bias in the samples, there are a number of advanced bootstrapping methods for determining appropriate limits for the CI, including the Bias Corrected and accelerated (BCa) method [32, 33], ABC, studentized, and double bootstrap [35]. The BCa bootstrap was selected due to its wide range of applicability [41], its relatively straightforward implementation, and its comparatively low computational requirements.

The BCa corrected $100(\alpha/2)^{th}$ percentile bootstrap index b_l and $100(1 - \alpha/2)^{th}$ percentile bootstrap index b_u, incorporating the small sample size correction (eq. 8), are given by eq. 9 and eq. 10 respectively, where $\lfloor x \rfloor$ is the integer floor and $\lceil x \rceil$ is the integer ceiling of x. Equation 11 is the bias correction factor

where Φ^{-1} is the inverse CDF of the normal distribution, θ is the calculated statistic, and θ_j^* is the jth bootstrap of the distribution of θ^*. Equation 13 is the acceleration factor derived from the original sample data. The CI_θ is $[\theta_{b_l}^{*,o}, \theta_{b_u}^{*,o}]$ where $\theta_b^{*,o}$ is the bth bootstrap of the ordered distribution of θ^*.

$$b_l = \left\lceil (B+1)\acute{p}\left(\Phi\left(z_0 + \frac{z_0 + \Phi^{-1}\left(\frac{\alpha}{2}\right)}{1 - acc \times \left(z_0 + \Phi^{-1}\left(\frac{\alpha}{2}\right)\right)}\right), n\right)\right\rceil \tag{9}$$

$$b_u = \left\lceil (B+1)\acute{p}\left(\Phi\left(z_0 + \frac{z_0 + \Phi^{-1}\left(1-\frac{\alpha}{2}\right)}{1 - acc \times \left(z_0 + \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right)}\right), n\right)\right\rceil \tag{10}$$

$$z_0 = \Phi^{-1}\left(\sum_{j=1}^{B} f(\theta_j^*, \theta)/B\right) \tag{11}$$

$$f(\theta_j^*, \theta) = \begin{cases} 1 \ if \ \theta_j^* < \theta \\ \frac{1}{2} \ if \ \theta_j^* = \theta \\ 0 \ if \ \theta_j^* > \theta \end{cases} \tag{12}$$

$$acc = \frac{\sum_{i=1}^{n}\left(\bar{\theta}_{(\cdot)} - \theta_{-i}\right)^3}{6\left[\sum_{i=1}^{n}\left(\bar{\theta}_{(\cdot)} - \theta_{-i}\right)^2\right]^{1.5}} \tag{13}$$

θ_{-i} is θ calculated using the original n simulations performed barring the ith simulation.

$$\bar{\theta}_{(\cdot)} = \frac{\sum_{i=1}^{n} \theta_{-i}}{n} \tag{14}$$

2.1.1 Bootstrapping the CI for the SD

If the TETs are normally distributed, the CI_{SD} can be determined using a chi-squared distribution with n-1 degrees of freedom [22]. However, in general the TETs will not be normally distributed and those theoretically derived CIs could be inappropriate. The CI_{SD} is therefore determined using bootstrapping.

For n simulations there are n TETs that are used to compute $SD^{(n)}$. A single bootstrap SD^* can be derived by randomly resampling, with replacement, those n TETs n times and calculating the SD of that bootstrap sample. This process is repeated B times, where $B \gg n$, and leads to a bootstrap *sampling* distribution of the SD^*. The CI of this ordered bootstrap *sampling* distribution of $SD^{*,o}$ approximates the CI for the true sampling distribution of $SD^{(n)}$. The BCa method is used to obtain the upper and lower limits of the CI.

The specific equations for calculating the BCa limits for the SD are obtained by substituting SD into eq. 9 to eq. 14 leading to eq. 15 to eq. 21.

$$z_{0,SD} = \Phi^{-1}\left(\sum_{j=1}^{B} f(SD_j^*, SD^{(n)})/B\right) \tag{15}$$

$$acc_{SD} = \frac{\sum_{i=1}^{n}\left(\overline{SD}_{(\cdot)} - SD_{-i}\right)^3}{6\left[\sum_{i=1}^{n}\left(\overline{SD}_{(\cdot)} - SD_{-i}\right)^2\right]^{1.5}} \tag{16}$$

$$SD_{-i} = \sqrt{\frac{\sum_{k=1}^{i-1}(TET_k - MT_{-i})^2 + \sum_{k=i+1}^{n}(TET_k - MT_{-i})^2}{n-2}} \tag{17}$$

$$MT_{-i} = \frac{\sum_{l=1}^{i-1} TET_l + \sum_{l=i+1}^{n} TET_l}{n-1} \tag{18}$$

$$\overline{SD}_{(\cdot)} = \frac{\sum_{i=1}^{n} SD_{-i}}{n} \tag{19}$$

$$b_{l,SD} = \left\lfloor (B+1)\acute{p}\left(\Phi\left(z_{0,SD} + \frac{z_{0,SD} + \Phi^{-1}\left(\frac{\alpha}{2}\right)}{1 - acc_{SD} \times \left(z_{0,SD} + \Phi^{-1}\left(\frac{\alpha}{2}\right)\right)}\right), n\right)\right\rfloor \tag{20}$$

$$b_{u,SD} = \left\lfloor (B+1)\acute{p}\left(\Phi\left(z_{0,SD} + \frac{z_{0,SD} + \Phi^{-1}\left(1-\frac{\alpha}{2}\right)}{1 - acc_{SD} \times \left(z_{0,SD} + \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right)}\right), n\right)\right\rfloor \tag{21}$$

The CI_{SD} is $[SD_{b_{l,SD}}^{*,o}, SD_{b_{u,SD}}^{*,o}]$ where the bootstrap index $b_{l,SD}$ is given by eq. 20 and $b_{u,SD}$ is given by eq. 21.

2.2 Functional Analysis

The final area of mathematics used in this paper to define CIs for the **AC** is functional analysis, a branch of mathematics used to compare vectors or times series data. An egress curve is a vector of agent egress times. Unlike a scalar quantity such as the TET where the difference between two TETs is merely a difference in magnitude, it is less obvious how to express the difference between vectors. The difference between the curves can be expressed using a set of FAMs.

The FAMs compare one curve (**a**) against another curve (**b**), using different criteria, that returns a scaler value that meaningfully represents the similarity (or difference) between the two curves. The egress curve $\mathbf{a} = \{a_1, a_k, ..., a_m\}$, where a_k is the time of the kth agent to exit and a_m is the exit time of the last (m^{th}) agent to exit; egress curve **b** is similarly defined. The three FAMs used here are the Euclidean Relative Distance (ERD), the Euclidean Projection Coefficient (EPC), and the Secant Cosine (SC). These FAMs have been previously described in fire [30] and evacuation contexts [16,17,19,29] so only concise details, suitable for egress curves, are given here.

The ERD (eq. 22) represents the "relative distance" between two curves and will be zero when the curves are identical and greater than zero if the curves are different. ERD varies between zero and infinity with values nearer to zero implying a closer "distance" between the curves. In the appendix it is shown that the ERD of two egress curves can be compared to normalised absolute difference (NAD), i.e. $|TET_a - TET_b| \div TET_b$, between two TETs of those egress curves. If the ERD is greater than the NAD, then it indicates there is more relative difference in the egress curves compared to the TETs; this suggests there may be more variability in an earlier phase of the evacuation. If the ERD is less than the NAD, then there is less relative difference in the curve compared to the TETs; this suggests that there may be more variability at the end of the evacuation.

$$ERD(\mathbf{a},\mathbf{b}) = \sqrt{\frac{\sum_{k=1}^{m}(a_k - b_k)^2}{\sum_{k=1}^{m} b_k^2}} \tag{22}$$

The EPC (eq. 23) is the value needed to project one curve onto another curve. This would be one when the curves can be exactly "projected" on each other, which will be the case when the curves are the same and varies about one when not exactly projected. EPC varies between zero and infinity, assuming $a_k \geq 0$ and $b_k \geq 0$ $\forall k$, with values nearer to one implying a closer "projection" similarity between the curves.

$$EPC(\mathbf{a},\mathbf{b}) = \frac{\sum_{k=1}^{m} a_k b_k}{\sum_{k=1}^{m} b_k^2} \tag{23}$$

The SC (eq. 24, where $\Delta x_{k-s} = x_k - x_{k-s}$ and s is the step-size ≥ 1) compares the "shape" of two curves. The step-size s is used to vary the smoothing of the curve, where a step size of one implies no smoothing

11

and the smoothing effect increases as s increases. SC equates to one when the curves have the same "shape", which will be the case when the curves are the same. SC varies between zero and one when the egress curves have different "shapes" with values nearer to one implying a closer similarity in shape. Egress curves monotonically increase, thus $\Delta a_{k-s} \geq 0$ and $\Delta b_{k-s} \geq 0 \; \forall k$, implying that $SC(\mathbf{a}, \mathbf{b}) \geq 0$. The Cauchy–Schwarz [42] inequality (eq. 25) implies that $SC(\mathbf{a}, \mathbf{b}) \leq 1$.

$$SC(\mathbf{a},\mathbf{b}) = \frac{\sum_{k=1+s}^{m} \Delta a_{k-s} \Delta b_{k-s}}{\sqrt{\sum_{k=1+s}^{m} (\Delta a_{i-s})^2 \sum_{k=1+s}^{m} (\Delta b_{k-s})^2}} \tag{24}$$

$$\left(\sum_{k=1+s}^{m} \Delta a_{k-s} \Delta b_{k-s}\right)^2 \leq \sum_{k=1+s}^{m} (\Delta a_{k-s})^2 \sum_{k=1+s}^{m} (\Delta b_{k-s})^2 \tag{25}$$

3 Development of FAM-based CIs for the **AC**

The average egress curve $\mathbf{AC}^{(n)} = \{ac_1, ac_k, ..., ac_m\}$ is calculated by averaging all n generated egress curves $\mathbf{c}_i$ (eq. 26 shows the calculation for the kth component of $\mathbf{AC}^{(n)}$) from the simulations performed. A bootstrap average curve $\mathbf{AC}^*$ is derived by randomly resampling, with replacement, those n generated curves n times, and then averaging. This process is repeated B times and leads to a bootstrap *sampling* distribution of the average egress curve. Unfortunately, it is difficult to directly assign a CI to the sampling distribution of curves. However, it is possible to create proxies that represent the precision of the **AC** indirectly. This can be achieved by transforming the **AC** sampling distribution of curves into usable sampling distributions of point values, via the FAMs, that can be assigned CIs. Each bootstrapped average curve $\mathbf{AC}^*$ is compared to the average egress curve $\mathbf{AC}^{(n)}$ with the FAMs to create sampling distributions of those measures, i.e. $ERD^*(\mathbf{AC}^*, \mathbf{AC}^{(n)})$, $EPC^*(\mathbf{AC}^*, \mathbf{AC}^{(n)})$, and $SC^*(\mathbf{AC}^*, \mathbf{AC}^{(n)})$. This approximates the (generally) unknowable sampling distributions of all possible sample average egress curves $\mathbf{AC}^{(n)}$ compared to the population (or expected) average egress curve $\mathbf{AC}^{(\infty)}$ using the FAMs, i.e., $ERD(\mathbf{AC}^{(n)}, \mathbf{AC}^{(\infty)})$, $EPC(\mathbf{AC}^{(n)}, \mathbf{AC}^{(\infty)})$, and $SC(\mathbf{AC}^{(n)}, \mathbf{AC}^{(\infty)})$. The CIs of the FAM-based sampling distributions represent the precision of the **AC**.

$$ac_k = \frac{\sum_{i=1}^{n} c_{i,k}}{n} \tag{26}$$

The ERD sampling distribution can be represented by a semi-infinite distribution between zero to infinity as the ERD values are greater than or equal to zero. It is anticipated that the distribution would be concentrated near zero with the frequency tending to zero as the bootstrapped ERDs tend to infinity. When $\mathbf{AC}^{(\infty)}$ is unknown the best estimate of $ERD^*(\mathbf{AC}^{(\infty)}, \mathbf{AC}^{(n)})$ is zero as the best estimate of $\mathbf{AC}^{(\infty)}$ is $\mathbf{AC}^{(n)}$. The 0th percentile of the $ERD^{*,o}$ sampling distribution is zero. The CI is obtained by using the 0th and $100(1 - \alpha)^{th}$ percentiles of the ordered bootstrapped values as the lower and upper limits for a confidence level of $100(1 - \alpha)\%$. The small sample size correction (eq. 8) is applied to the $100(1 - \alpha)^{th}$ percentile; this is achieved by modelling the bootstrapped ERD measures using a folded normal distribution from zero to infinity with the (unfolded and folded) mode at zero. The corrected $100(1 - \alpha)^{th}$ percentile bootstrap index b_{ERD} of the ordered bootstrapped $ERD^{*,o}$ values is given by eq. 27 where eq. 28 and eq. 29 map the percentile from the folded distribution to the full distribution and back again. The CI_{ERD} is $[0, ERD^{*,o}_{b_{ERD}}]$.

$$b_{ERD}(\alpha) = \left\lceil (B + 1) p_{folded}\big(\acute{p}(p_{full}(1 - \alpha), n)\big) \right\rceil \tag{27}$$

$$p_{full}(p_{folded}) = \frac{p_{folded} + 1}{2} \tag{28}$$

$$p_{folded}(p_{full}) = 2p_{full} - 1 \tag{29}$$

The SC sampling distribution can be represented by a distribution between zero and one where it is expected that the values will be concentrated near one and the frequency will tend to zero as the SC values tend to zero. The best estimate of $SC(\mathbf{AC}^{(\infty)}, \mathbf{AC}^{(n)})$ is one and the 100th percentile of the SC sampling distribution is one. The CI, with a confidence level of $100(1 - \alpha)\%$, is given by the limits of

the corrected $(100\alpha)^{th}$ and 100th percentile of the ordered bootstrapped $SC^{*,o}$ measures. The corrected $100\alpha^{th}$ percentile bootstrap index b_{SC} is modelled, in a similar fashion to b_{ERD}, using eq. 30 where eq. 31 and eq. 32 map the percentile from the folded distribution to the full distribution and back again. The CI_{SC} is $[SC^{*,o}_{b_{SC}}, 1]$.

$$b_{SC}(\alpha) = \lfloor (B+1)p_{folded}(\acute{p}(p_{full}(\alpha), n)) \rfloor \tag{30}$$

$$p_{full}(p_{folded}) = \frac{p_{folded}}{2} \tag{31}$$

$$p_{folded}(p_{full}) = 2p_{full} \tag{32}$$

The EPC sampling distribution will vary between zero and infinity where it is expected that the values will be concentrated about one and the frequency tends to zero as EPC tends to either zero or infinity. This can be represented by a two-tailed distribution and the CI of the ordered bootstrapped EPC measures is calculated using the BCa method (section 2.4). The best estimate of $EPC(\mathbf{AC}^{(\infty)}, \mathbf{AC}^{(n)})$ is one. The specific equations for calculating the BCa limits for the EPC are obtained by substituting EPC into eq. 9 to eq. 14 leading to eq. 33 to eq. 37.

$$z_{0,EPC} = \Phi^{-1}\left(\sum_{j=1}^{B} f\left(EPC\left(\mathbf{AC}^*_j, \mathbf{AC}^{(n)}\right), 1\right)/B\right) \tag{33}$$

$$acc_{EPC} = \frac{\sum_{i=1}^{n}\left(\overline{EPC}_{(\cdot)} - EPC\left(\mathbf{AC}^{(n)}_{-i}, \mathbf{AC}^{(n)}\right)\right)^3}{6\left[\sum_{i=1}^{n}\left(\overline{EPC}_{(\cdot)} - EPC\left(\mathbf{AC}^{(n)}_{-i}, \mathbf{AC}^{(n)}\right)\right)^2\right]^{1.5}} \tag{34}$$

Where $\mathbf{AC}^{(n)}_{-i}$ is an average egress curve calculated using the original egress curves barring the egress curve from the ith simulation.

$$\overline{EPC}_{(\cdot)} = \frac{\sum_{i=1}^{n} EPC\left(\mathbf{AC}^n_{-i}, \mathbf{AC}^n\right)}{n} \tag{35}$$

$$b_{l,EPC}(\alpha) = \left\lfloor (B+1)\acute{p}\left(\Phi\left(z_{0,EPC} + \frac{z_{0,EPC}+\Phi^{-1}\left(\frac{\alpha}{2}\right)}{1-acc_{EPC}\times\left(z_{0,EPC}+\Phi^{-1}\left(\frac{\alpha}{2}\right)\right)}\right), n\right) \right\rfloor \tag{36}$$

$$b_{u,EPC}(\alpha) = \left\lfloor (B+1)\acute{p}\left(\Phi\left(z_{0,EPC} + \frac{z_{0,EPC}+\Phi^{-1}\left(1-\frac{\alpha}{2}\right)}{1-acc_{EPC}\times\left(z_{0,EPC}+\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right)}\right), n\right) \right\rfloor \tag{37}$$

The CI_{EPC} is given by $[EPC^{*,o}_{b_{l,EPC}}, EPC^{*,o}_{b_{u,EPC}}]$ where $b_{l,EPC}$ and $b_{u,EPC}$ are given by eq. 36 and eq. 37 respectively.

3.1 Overall confidence level for average egress curve

The precision of the **AC** is represented by three FAM-based CIs. Each of these CIs has an individual CL of $100(1 - \alpha)\%$. These three CIs can be considered to form a triple-CI that has a more complex coverage probability than a single CI as it is possible that the $\mathbf{AC}^{(\infty)}$ may lie within zero, one, two, or three of the CIs (see Fig. 4).
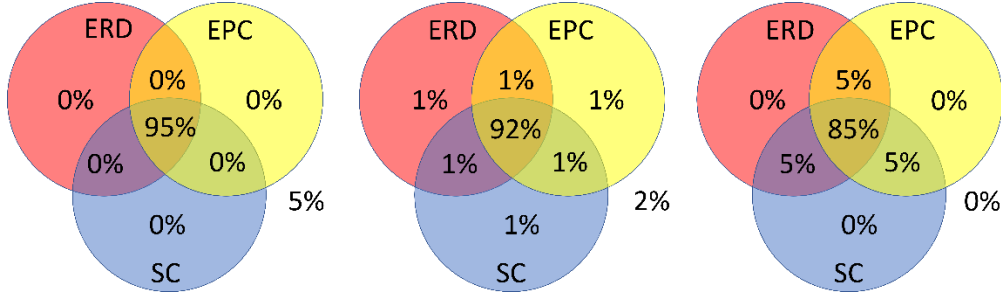
Fig. 4 Three potential triple-CI coverage probabilities where the individual ERD, EPC, and SC CLs are 95%. The value outside of the circles represents the case when the $\mathbf{AC}^{(\infty)}$ does not lie in any CI. The left-hand case is when the overall CL, $P(\text{ERD} \cap \text{EPC} \cap \text{SC})$, is equivalent to the individual CLs. The middle case is a mixture of CI occupancies. The right-hand case is where the overall CL is the minimum possible value.

The overall confidence level is equivalent to $P(\text{ERD} \cap \text{EPC} \cap \text{SC})$, i.e. all FAM-based CIs are simultaneously satisfied, and is likely to be less than the individual CL. Although it is easy to control the individual CLs, what is sometimes required is the ability to set the overall CL. This is a well-known problem with multiple CIs and could be addressed by applying the Bonferroni correction to the individual CIs [43]. If the overall CL is $100(1-\alpha_{\text{overall}})\%$ then the confidence level required for each of the individual CIs is given be eq. 38 where N_{CI} is the number of CIs (three for the average egress curve).

$$CL_{\text{Bonferroni}} = 100(1 - \alpha_{\text{Bonferroni}})\% = 100\left(1 - \frac{\alpha_{\text{overall}}}{N_{\text{CI}}}\right)\% \tag{38}$$

However, this leads to a conservative estimate of the individual CLs, as the correction assumes the worst-case multi-CI probability coverage, e.g. the right-hand Venn diagram of Fig. 4, and therefore the CIs are potentially much wider than necessary. When three CIs are used, an overall 95% CL corresponds to each CI having an individual 98.3% CL. Fortunately, as the $ERD^*(\mathbf{AC}^*, \mathbf{AC}^{(n)})$, $EPC^*(\mathbf{AC}^*, \mathbf{AC}^{(n)})$, and $SC^*(\mathbf{AC}^*, \mathbf{AC}^{(n)})$ sampling distributions are available, for the bootstrapped egress curves, it is possible to determine a more accurate overall confidence level. This is achieved by adjusting the individual confidence level of the CIs, using the bisection algorithm below, until the number of $\mathbf{AC}^*$ bootstrap curves that are simultaneously within all three FAM-based CIs matches the required overall CL.

1. The required number (R_{AC}) (eq. 39) of $\mathbf{AC}^*$ curves that needs to be contained simultaneously by all three CIs, for an overall CL of $100(1-\alpha_{\text{overall}})\%$, is calculated by taking the maximum of the number of $\mathbf{AC}^*$ that need to be held individually by CI_{ERD} (eq. 40), CI_{EPC} (eq. 41), and CI_{SC} (eq. 42) at the overall CL.

$$R_{\text{AC}} = \max(NB_{\text{ERD}}, NB_{\text{EPC}}, NB_{\text{SC}}) \tag{39}$$

$$NB_{\text{ERD}} = b_{\text{ERD}}(\alpha_{\text{overall}}) \tag{40}$$

$$NB_{\text{EPC}} = b_{\text{u,EPC}}(\alpha_{\text{overall}}) - b_{\text{l,EPC}}(\alpha_{\text{overall}}) + 1 \tag{41}$$

$$NB_{\text{SC}} = B - b_{\text{SC}}(\alpha_{\text{overall}}) + 1 \tag{42}$$

2. The lower individual confidence level (*LICL*) and upper individual confidence level (*UICL*) for bisection algorithm are set to eq. 43 and eq. 44 (the Bonferroni correction) respectively. The actual individual CL, that satisfies the overall CL, lies between these two limits.

$$LICL = 100(1 - \alpha_{\text{overall}})\% \tag{43}$$

$$UICL = 100(1 - \alpha_{\text{overall}}/3)\% \tag{44}$$

14

3. The bisection is now performed by setting the middle individual confidence level (*MICL*) eq. 45.

$$MICL = \frac{LICL + UICL}{2} \tag{45}$$

4. The CI limits based on *MICL*, $100(1-\alpha_{MICL})\%$, are calculated for the ERD, EPC and SC sampling distributions, i.e. $ERD^{*,o}_{b_{ERD}(\alpha_{MICL})}$ (eq. 27), $EPC^{*,o}_{b_{l,EPC}(\alpha_{MICL})}$ (eq. 36) and $EPC^{*,o}_{b_{u,EPC}(\alpha_{MICL})}$ (eq. 37), and $SC^{*,o}_{b_{SC}(\alpha_{MICL})}$ (eq. 30).

5. Count the number (N_{AC}) of $\mathbf{AC}^*$ curves that are simultaneously contained by all three CIs with MICL. This is achieved using the following pseudo code. Note that the FAMs of $\mathbf{AC}^*_j$ compared to $\mathbf{AC}^{(n)}$ are ERD^*_j, EPC^*_j and, SC^*_j.

> Set N_{AC} to zero
> DO $j = 1$ to B
> > IF $\left(ERD^*_j \leq ERD^{*,o}_{b_{ERD}(\alpha_{MICL})}\right)$ AND $\left(SC^*_j \geq SC^{*,o}_{b_{SC}(\alpha_{MICL})}\right)$
> > AND $\left(EPC^{*,o}_{b_{l,EPC}(\alpha_{MICL})} \leq EPC^*_j \leq EPC^{*,o}_{b_{u,EPC}(\alpha_{MICL})}\right)$ THEN $N_{AC} = N_{AC} + 1$
> > ENDIF
> ENDDO

6. If ($N_{AC} > R_{AC}$) then set *UICL* to *MICL*. Otherwise set *LICL* = *MICL*.

7. If the difference between *UICL* and *LICL* is less than the bisection tolerance, 0.1% is suggested, then the overall CL convergence has been achieved (continue to step 8). Otherwise, the overall CL convergence has not been achieved (go to step 3).

8. Calculate the final CIs using the *UICL* by substituting α_{UICL} into $ERD^{*,o}_{b_{ERD}(\alpha_{UICL})}$ (eq. 27), $EPC^{*,o}_{b_{l,EPC}(\alpha_{UICL})}$ (eq. 36) and $EPC^{*,o}_{b_{u,EPC}(\alpha_{UICL})}$ (eq. 37), and $SC^{*,o}_{b_{SC}(\alpha_{UICL})}$ (eq. 30) for an overall CL of $100(1-\alpha_{overall})\%$.

4 CI convergence scheme

Now that CIs have been developed for the statistics it is possible to determine convergence w.r.t. the estimate of the population parameters. The CI_{MT} is calculated using eq. 3, the CI_{SD} is calculated using bootstrapping (section 2.1.1), and the FAM-based CIs of the **AC** are calculated using the methodology described in section 3. The width of a CI narrows (converges) with increasing sample size, e.g. eq. 4. This enables a convergent approach based on incremental testing where the number of trials can be progressively increased until the width of the CI is sufficiently reduced that it is less than the specified precision tolerance. The convergence of the MT, the SD, and the FAMs are determined by comparing the normalised width (*W*) of the CIs against a specified tolerance (*Tol*) (eq. 46 to eq. 50) for a $100(1-\alpha)\%$ CL.

$$W_{MT} = \frac{CI_{u,MT} - CI_{l,MT}}{MT} = \frac{2 . T^{-1}\left(1 - \frac{\alpha}{2}, n-1\right) . SD}{MT \sqrt{n}} < Tol_{MT} \tag{46}$$

$$W_{SD} = \frac{CI_{u,SD} - CI_{l,SD}}{SD^{(n)}} = \frac{SD^{*,o}_{b_{u,SD}(\alpha)} - SD^{*,o}_{b_{l,SD}(\alpha)}}{SD^{(n)}} < Tol_{SD} \tag{47}$$

$$W_{ERD} = CI_{u,ERD} - CI_{0^{th},ERD} = ERD^{*,o}_{b_{ERD}(\alpha)} < Tol_{ERD} \tag{48}$$

$$W_{\text{EPC}} = CI_{\text{u, EPC}} - CI_{\text{l, EPC}} = EPC^{*,o}_{b_{\text{u, EPC}}(\alpha)} - EPC^{*,o}_{b_{\text{l,EPC}}(\alpha)} < Tol_{\text{EPC}} \quad (49)$$

$$W_{\text{SC}} = CI_{100^{\text{th}}, \text{SC}} - CI_{\text{l, SC}} = 1 - SC^{*,o}_{b_{\text{SC}}(\alpha)} < Tol_{\text{SC}} \quad (50)$$
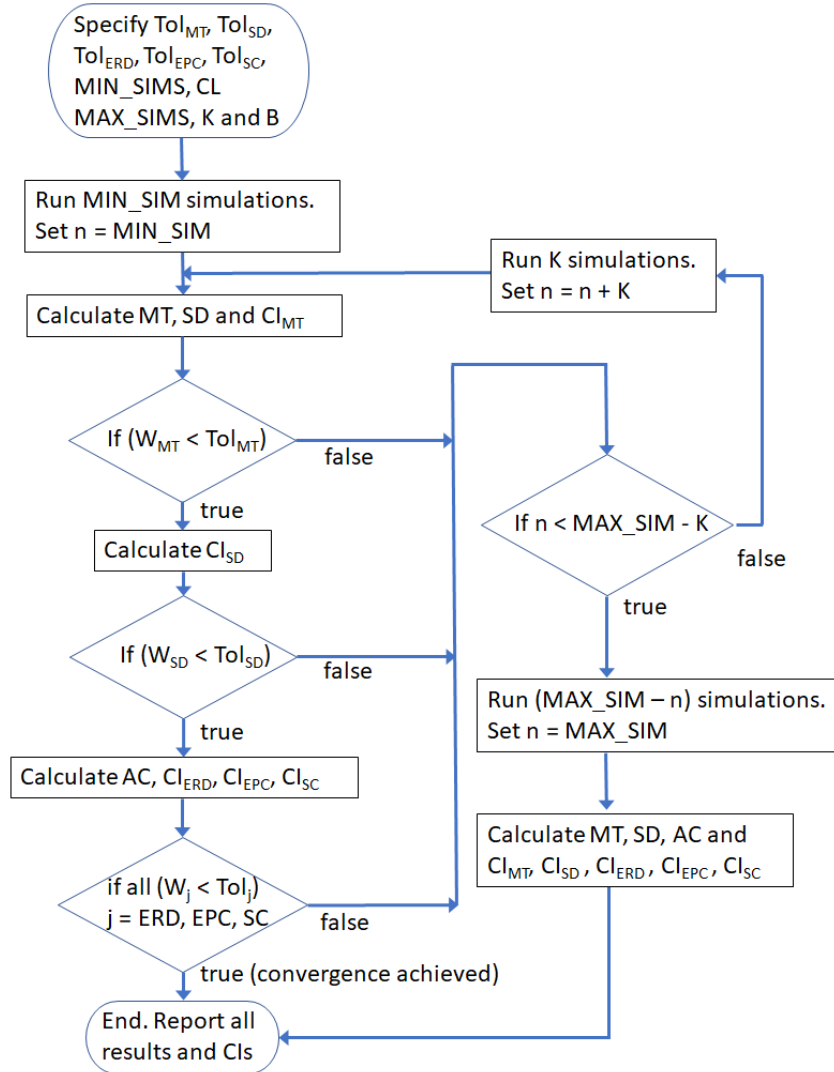

Fig. 5 – CI-based convergence algorithm using the MT, SD, and FAMs of the **AC**.

The convergence algorithm is depicted as a flowchart in Fig. 5. The user needs to specify the above tolerances (Tol_{MT}, Tol_{SD}, Tol_{ERD}, Tol_{EPC}, and Tol_{SC}) as well as the minimum number of simulations (MIN_SIM), the maximum number of simulations (MAX_SIM), the number of bootstrap-resamples (B), the CL, and the number of simulations performed between testing (K). Convergence is determined when all the CI widths are less than their specified tolerances. Convergence may not be achieved within the MAX_SIM simulations, in that case the MAX_SIM simulations will be performed with the final CIs reported. If the user has no specific convergence requirements, then a set of values for the convergence attributes are suggested in Table 6. The tolerances are discussed in section 6.

Table 6 – Suggested values for the attributes of the CI convergence algorithm.

Attribute	MIN_SIM	CL	K	B	Tol_{MT}	Tol_{SD}	Tol_{ERD}	Tol_{EPC}	Tol_{SC}
Value	40+ [31]	95%	1	2000+ [32]	0.02	0.2	½Tol_{MT}	Tol_{MT}	½Tol_{MT}

5 Case Study

A theoretical case study model, similar to Ronchi et al. [17], was used to create multiple fictitious egress curves of 120 agents. The data is generated by pseudo-randomly [44] sampling a log-normal distribution with a mean of 12s and standard deviation of 13.4s ($\sqrt{180}$s). A set of 120 values are generated $\{l_1, ..., l_{120}\}$ with the egress time (t) of the kth agent given by eq. 51. Figure 6 depicts ten randomly generated egress curves using the case study model.

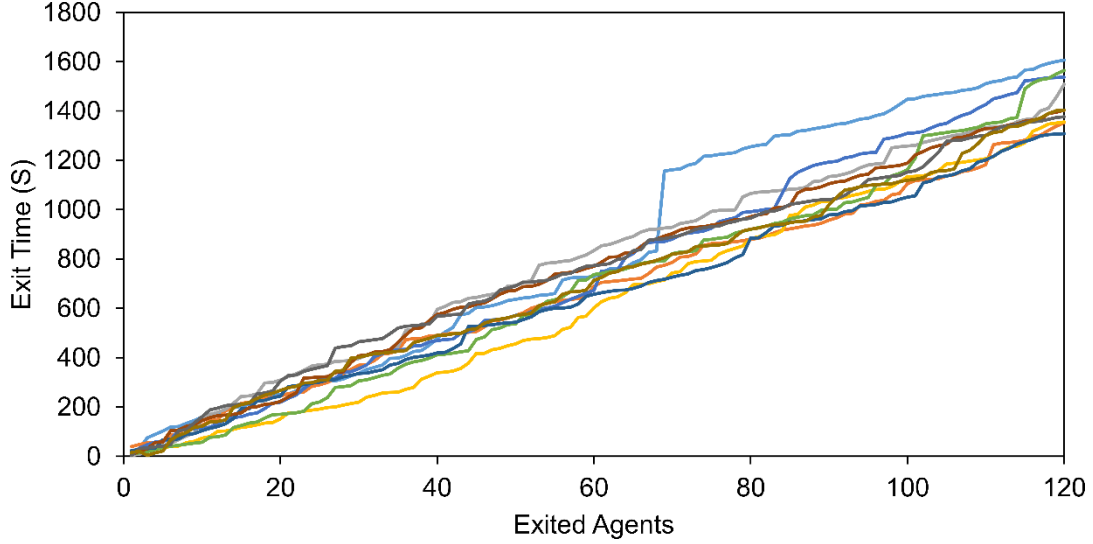$$t_k = \sum_{q=1}^{k} l_q \qquad (51)$$

Fig. 6 – Ten randomly generated fictitious egress curves for 120 agents from the case study model.

Using the case study model has two major advantages over using simulated or real data. Firstly, the generally unknowable *population* parameters are defined for the case study model so the methods can be tested and verified. The *population* MT$^{(\infty)}$ is 1440s (120×12s), as the mean of the sum is the sum of the means. The *population* SD$^{(\infty)}$ is 147.0s ($\sqrt{120}\sqrt{180}$s), as the variance of the sum is the sum of the variances and the standard deviation is the square root of the variance. The exit time of the kth agent of AC$^{(\infty)}$ is given by eq. 52.

$$t_k^{(\infty)} = k \times 12s \qquad (52)$$

Secondly, there is very little computational cost in generating this data allowing a vast number of curves to be generated in a short timeframe. Virtually all the computational cost of the case study is incurred from the bootstrapping process rather than curve generation. In practice, it is anticipated that the computational cost of egress simulations (curve generation) will dwarf the computational costs of the bootstrapping process. In this test case study, the step size (s) used for the SC (eq. 24) is one. This gives the most conservative value for the difference in shape between the curves.

The testing of the CIs is performed in three parts. The first part examines the behaviour of the CIs, within a single set of simulations, relative to the sample estimates of the statistics and the population parameters (section 5.1). The second part examines the average behaviour of the CIs over multiple sets of simulations including the measured coverage probability and average CI width for a parameter at a specified sample size (section 5.2). The third part examines the behaviour of the convergence scheme based on the CIs (section 5.3).

5.1 CI and error behaviour within a single set of simulations

The behaviour of the statistics (MT, SD, ERD, EPC, SC) and their CIs were examined over a set of two hundred simulations. This is an example set of simulations and the behaviour will not be the same every time but highlights the expected behaviour. The statistics are compared to their equivalent population parameters and normalised leading to the true errors (ε). The CIs are also normalised, in a similar fashion to ε, leading to normalised CI limits (v).

The normalised error ($\varepsilon_{\mathrm{MT}}$) and CI limits ($v_{\mathrm{l,MT}}, v_{\mathrm{u,MT}}$) for the MT are given by eq. 53, eq. 54, and eq. 55.

$$\varepsilon_{\mathrm{MT}} = \frac{MT^{(\infty)} - MT^{(n)}}{MT^{(n)}} \tag{53}$$

$$v_{\mathrm{l,MT}} = \frac{T^{-1}\left(\frac{\alpha}{2}, n-1\right) \times \frac{SD^{(n)}}{\sqrt{n}}}{MT^{(n)}} \tag{54}$$

$$v_{\mathrm{u,MT}} = \frac{T^{-1}\left(\left(1-\frac{\alpha}{2}\right), n-1\right) \times \frac{SD^{(n)}}{\sqrt{n}}}{MT^{(n)}} \tag{55}$$

The normalised error ($\varepsilon_{\mathrm{SD}}$) and CI limits ($v_{\mathrm{l,SD}}, v_{\mathrm{u,SD}}$) for the SD are given by eq. 56, eq. 57, and eq. 58.

$$\varepsilon_{\mathrm{SD}} = \frac{SD^{(\infty)} - SD^{(n)}}{SD^{(n)}} \tag{56}$$

$$v_{\mathrm{l,SD}} = \frac{SD_{b_{\mathrm{l,SD}}}^{*,o} - SD^{(n)}}{SD^{(n)}} \tag{57}$$

$$v_{\mathrm{u,SD}} = \frac{SD_{b_{\mathrm{u,SD}}}^{*,o} - SD^{(n)}}{SD^{(n)}} \tag{58}$$

The normalised error ($\varepsilon_{\mathrm{ERD}}$) and CI limit ($v_{\mathrm{ERD}}$) for the ERD are given by eq. 59 and eq. 60.

$$\varepsilon_{\mathrm{ERD}} = ERD\left(\mathbf{AC}^{(\infty)}, \mathbf{AC}^{(n)}\right) \tag{59}$$

$$v_{\mathrm{ERD}} = ERD_{b_{\mathrm{ERD}}}^{*,o} \tag{60}$$

The normalised error ($\varepsilon_{\mathrm{EPC}}$) and CI limits ($v_{\mathrm{l,EPC}}, v_{\mathrm{u,EPC}}$) for the EPC are given by eq. 61, eq. 62, and eq. 63.

$$\varepsilon_{\mathrm{EPC}} = EPC\left(\mathbf{AC}^{(\infty)}, \mathbf{AC}^{(n)}\right) - 1 \tag{61}$$

$$v_{\mathrm{l,EPC}} = EPC_{b_{\mathrm{l,EPC}}}^{*,o} - 1 \tag{62}$$

$$v_{\mathrm{u,EPC}} = EPC_{b_{\mathrm{u,EPC}}}^{*,o} - 1 \tag{63}$$

The normalised error ($\varepsilon_{\mathrm{SC}}$) and CI limit ($v_{\mathrm{SC}}$) for the SC are given by eq. 64 and eq. 65.

$$\varepsilon_{\mathrm{SC}} = 1 - SC\left(\mathbf{AC}^{(\infty)}, \mathbf{AC}^{(n)}\right) \tag{64}$$

$$v_{\mathrm{SC}} = 1 - SC_{b_{\mathrm{SC}}}^{*,o} \tag{65}$$

In Fig. 7 to Fig. 11 the normalised errors and CIs are plotted for MT (Fig. 7), SD (Fig. 8), ERD (Fig. 9), EPC (Fig. 10), and SC (Fig. 11).
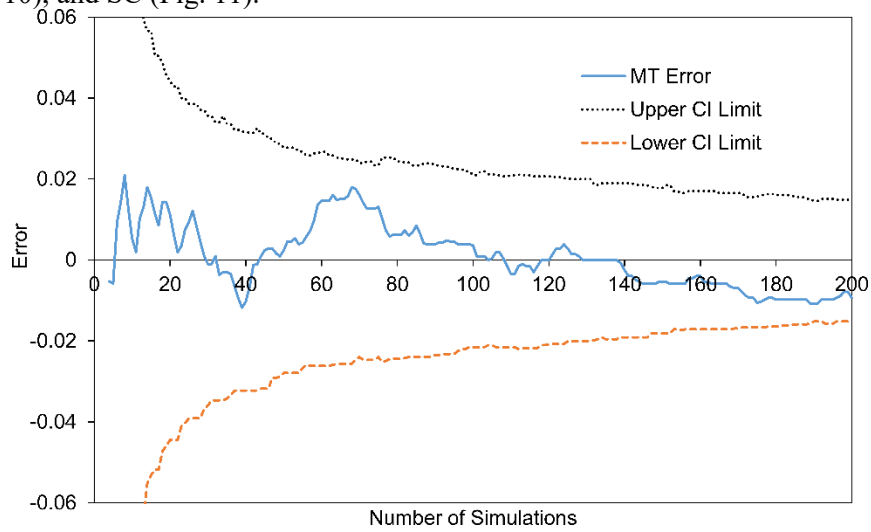

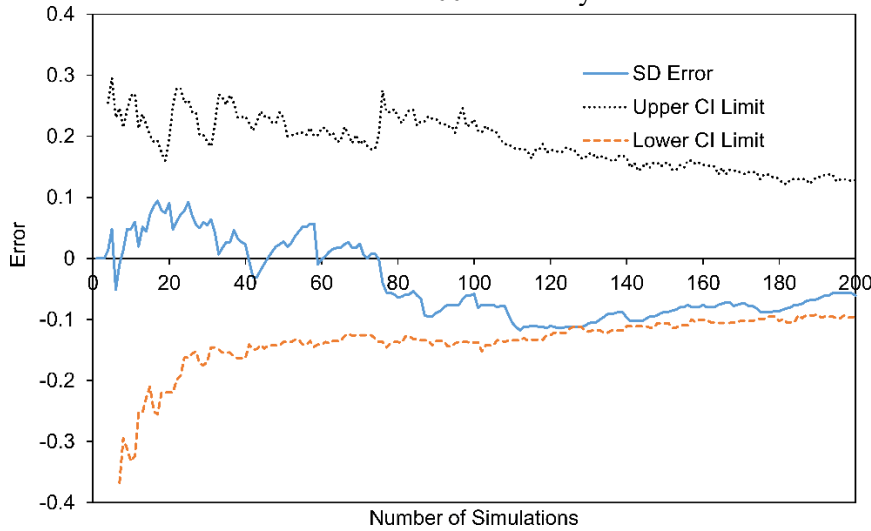Fig. 7 – MT Error and CI behaviour over a set of 200 case study simulations.

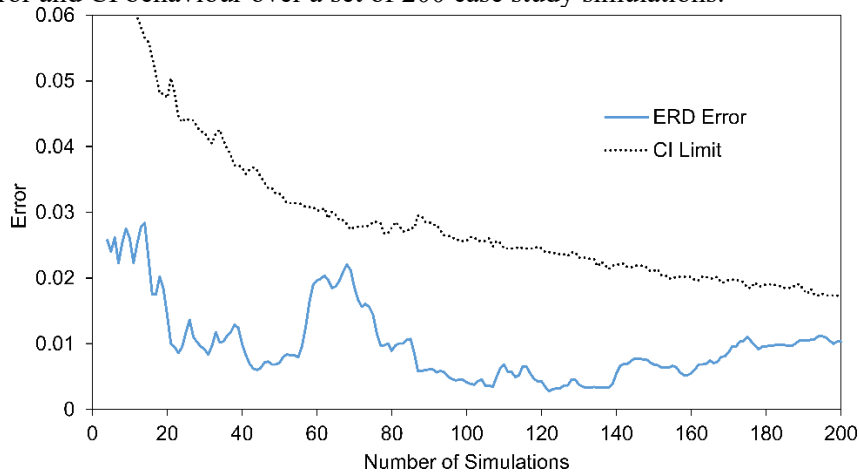Fig. 8 - SD Error and CI behaviour over a set of 200 case study simulations.

Fig. 9 - ERD Error and CI behaviour over a set of 200 case study simulations.
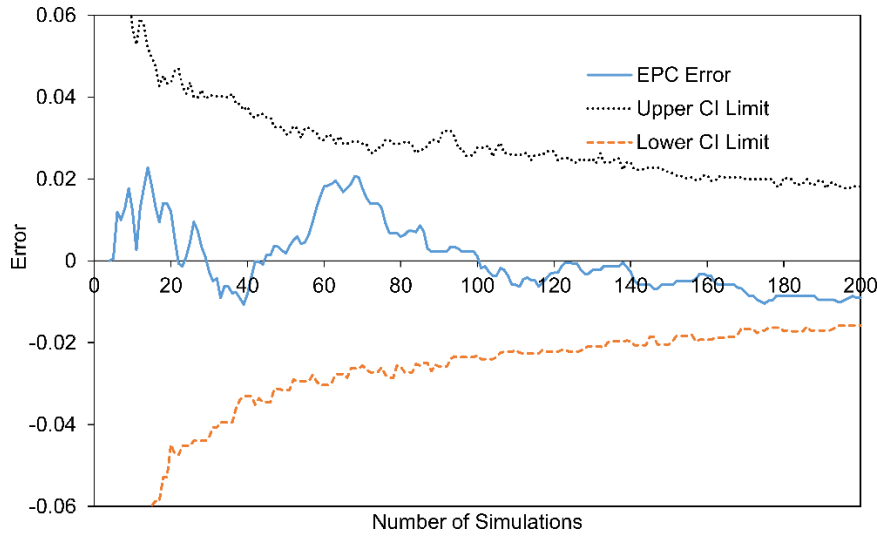
Fig. 10 – EPC Error and CI behaviour over a set of 200 case study simulations.
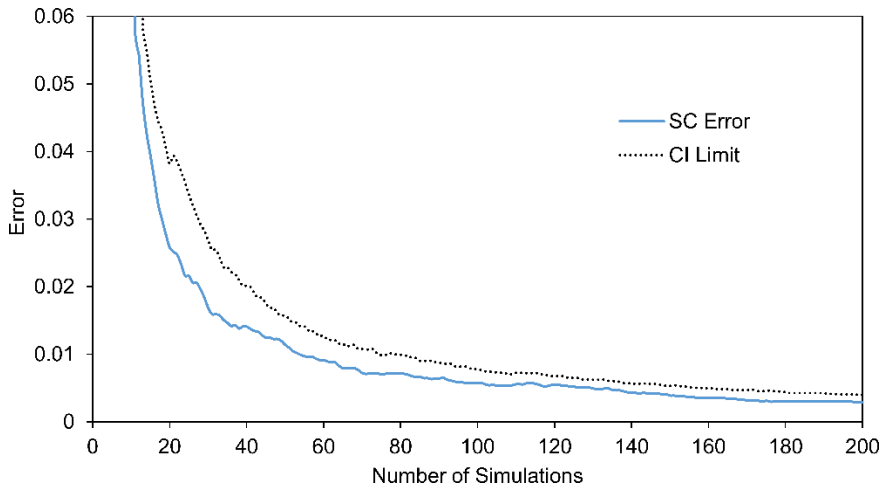

Fig. 11 – SC Error and CI behaviour over a set of 200 case study simulations.

In Fig. 7 to Fig. 11 it is apparent that for this set of simulations the CIs bounds the true error. This is expected as the CI is an estimate interval of the population parameter, meaning that the population parameter is unlikely to be outside of the CI. In Fig. 7, the SD has a significant error, ~0.1 (~10%), after 150 simulations, although this error is bounded by the CI. There is no obvious relationship between ε_{SD} and the other error measures.

There is an unexpectedly close similarity between the behaviour of ε_{MT} (Fig. 7) and the ε_{EPC} (Fig. 10), both errors are near zero at 50 and 100 simulations, have a similar local peak value, approximately 0.02, at 67 simulations, and have similar values of approximately -0.01 at 200 simulations. There is also a similarity between ε_{ERD} (Fig. 9) and $|\varepsilon_{MT}|$, both errors are near zero at 50 and 100 simulations, have a local peak value at 67 simulations, and have similar values of 0.01 at 200 simulations. Similarly, the CI limits for EPC and ERD have a similar trend and magnitude as the corresponding CI limits of MT. A brief algebraic analysis of ε_{EPC} and ε_{ERD} is given in the appendix. That analysis demonstrates that ε_{EPC} and ε_{ERD} are approximately equivalent to ε_{MT} under certain conditions. This is useful as the ERD and EPC tolerances, in the convergence scheme, can be related to the MT tolerance (see section 6). It is also shown, in the appendix, that if ε_{ERD} is greater than $|\varepsilon_{MT}|$ then there is more relative difference exhibited between the egress curves than the difference between the TETs of those egress curves.

The convergence behaviour of SC (Fig. 11) is somewhat different, and shows no obvious correlation, to the other statistics as the error and CI change comparatively smoothly with an increasing number of

simulations. The error reduces more rapidly than the other statistics with the normalised CI less than 0.01 after 100 simulations and less than 0.005 after 200 simulations. The convergence behaviour unexpectedly appears to be $1/n$ while the convergence of the other variables appears to be $1/\sqrt{n}$. Although the general behaviour is different to the other CIs the normalised CI_{SC} still bounds ε_{SC}.

5.2 Coverage Probability and average width of CIs

Ten thousand sets of simulations of size n were generated to estimate the actual coverage probability of the CIs and the average normalised widths (eq. 46 to eq. 50) of the CIs. This was repeated using eight sample sizes (10, 20, 30, 40, 100, 400, 1000, and 4000). The coverage probability is the percentage of CIs that "contain" the population parameter, and ideally this should match the confidence level of the CI. For the MT and SD, the coverage probability is determined by counting the number of times the population value lies within the CI. For the average egress curve, the FAMs are used to compare the sample average egress curve to the population average egress curve, e.g., $ERD(\mathbf{AC}^{(\infty)}, \mathbf{AC}^{(n)})$, $EPC(\mathbf{AC}^{(\infty)}, \mathbf{AC}^{(n)})$, and $SC(\mathbf{AC}^{(\infty)}, \mathbf{AC}^{(n)})$, and those values were checked to see if they lie within the CI or not.

In Table 7, the coverage probability of the CIs is generally good, the CI_{MT} and CI_{ERD} coverage probability nearly perfectly matching the confidence level within the error band. The CI_{EPC} coverage probability is within the error band for sample sizes of 30 or more; even at a sample size of 10 the discrepancy is not too great. However, the CI_{SC} coverage probability is unexpectedly higher than the nominal confidence level until a sample size of 4,000 is reached. This implies that the CI_{SC} may not be as narrow as they could be, but at least the CI is a conservative estimate. The CI_{SD} has the worst coverage probability behaviour of all the statistics but is still reasonable (>90%) with sample sizes over 20. Even for a small sample size of 10 the coverage probability is 85.2% and is perhaps better than no error estimate. The CI_{SD} coverage probability tends to the nominal CL with increasing sample size, but even for small sample sizes the method is usable. The coverage probability could perhaps be improved for the CI_{SD} using a double bootstrap [35], but this will result in a significant additional computational cost.

Table 7 – CI coverage probability of MT, SD, ERD, EPC, and SC with individual 95% CLs for a range of sample sizes over 10,000 repetitions.

Statistic	Error band	Number of Simulations							
		10	20	30	40	100	400	1000	4000
MT (95%)	±0.4%	94.5%	94.8%	95.0%	94.6%	95.1%	95.2%	95.3%	94.8%
SD (95%)	±0.4%	85.2%	91.7%	92.4%	93.0%	93.3%	94.6%	94.7%	94.9%
ERD (95%)	±0.4%	95.0%	94.7%	95.2%	95.6%	95.3%	94.9%	94.8%	95.0%
EPC (95%)	±0.4%	94.2%	94.4%	95.1%	95.2%	95.1%	94.8%	95.1%	94.9%
SC (95%)	±0.4%	100%	100%	99.9%	99.8%	99.3%	97.7%	96.1%	95.4%

In Table 8, the widths of the MT, SD, ERD, and EPC CIs are approximately inversely proportional to the square root of the number of simulations performed. This is a typical relationship between the CI width and the number of simulations performed and is similarly observed for CI widths of other parameters [13]. The SC's width has a different convergence behaviour and is approximately inversely proportional to the number of simulations performed.

Table 8 - Average normalised CI widths for MT, SD, ERD, EPC, and SC for eight sample sizes over 10,000 repetitions.

Width of normalised CI	Number of simulations							
	10	20	30	40	100	400	1000	4000
W_{MT} (95%) (eq. 46)	0.14	0.094	0.076	0.065	0.04	0.02	0.013	0.0063
W_{SD} (95%) (eq. 47)	0.84	0.68	0.57	0.49	0.31	0.16	0.098	0.049
W_{ERD} (95%) (eq. 48)	0.084	0.053	0.043	0.037	0.023	0.011	0.0072	0.0036
W_{EPC} (95%) (eq. 49)	0.16	0.10	0.083	0.071	0.044	0.022	0.014	0.0069
W_{SC} (95%) (eq. 50)	0.10	0.047	0.03	0.022	0.0084	0.002	0.00077	0.00019

The normalised width of the SD is notable as it is much wider than the other statistics. For example, after 1,000 simulations the W_{SD} is still about 10% of the SD, approximately ±5% error. For 40 simulations, the W_{SD} is 0.49 which is equivalent to a [-19%, +30%] error at a 95% CL. For 10 simulations, the W_{SD} is 0.84 is equivalent to a [-24%, +60%] error. If an end user requires a precise estimate of the SD, then a high number of simulations will generally be required.

When the multi-CI correction scheme (section 3.1) is used it can be seen in Table 9 that the average individual CL (97.6%) is significantly lower than the conservative Bonferroni correction (98.3%), but the overall coverage probability is still above 95%. It should be noted that the SC coverage probability is much greater than its CL, thus significantly contributing to the overall coverage probability being over 95% below 4,000 simulations. However, with the 4,000-simulation sized sample the overall coverage probability is still above 95% when the SC coverage probability is in line with the expectation of the computed CL.

Table 9 - CI coverage probability for a range of sample sizes over 10,000 repetitions when the multi-CI correction scheme is used for **AC** with an overall CL of 95%.

	Error band	Number of Simulations							
		10	20	30	40	100	400	1000	4000
Average individual CL	±0.1%	97.1%	97.3%	97.4%	97.5%	97.5%	97.6%	97.6%	97.5%
ERD	±0.3%	97.2%	97.4%	97.2%	97.4%	97.5%	97.5%	97.5%	97.5%
EPC	±0.3%	96.5%	97.1%	97.0%	97.5%	97.5%	97.4%	97.5%	97.5%
SC	±0.3%	100%	100%	100%	100%	99.9%	99.1%	98.4%	97.6%
Overall coverage	±0.4%	95.9%	96.6%	96.6%	96.7%	97.0%	96.2%	95.7%	95.2%

In Table 10 the corrected widths for an overall nominal 95% CL are given. These widths are slightly wider (~14%) than the widths in Table 8. For 40 simulations, the CI_{ERD} is [0, 0.042], CI_{EPC} is [0.958, 1.042], and CI_{SC} is [0.975, 1] with an overall confidence level of 95%. These could be reported as errors with the ERD error of 4.2%, the EPC error of ±4.2% and the SC error of 2.5%.

Table 10 - Average multi-CI corrected normalised CI widths for **AC** at various sample sizes over 10,000 repetitions.

normalised CI width	Number of simulations							
	10	20	30	40	100	400	1000	4000
W_{ERD}	0.094	0.061	0.049	0.042	0.026	0.013	0.0082	0.0041
W_{EPC}	0.18	0.12	0.096	0.082	0.051	0.025	0.016	0.0079
W_{SC}	0.12	0.054	0.034	0.025	0.0091	0.0021	0.00081	0.00020

5.3 Convergence Testing

A sample of simulations is generated until convergence was achieved for the tested statistic at the stated tolerance. A minimum of 40 simulations [41] were performed before testing for convergence. This was

repeated 10,000 times for each measure to estimate the actual coverage probability of the CI. For this case study each convergence criteria were tested in isolation, i.e., convergence is based solely on the tested statistic of interest. This is a reasonable way to test the convergence as ultimately convergence will be dependent on a single variable as the other variables must be individually converged before the final convergence is determined.

The results in Table 11 illustrate that the methodology can generate CIs that have an actual coverage close to the nominal 95% confidence level. The theory of CIs is based on a variable CI limits with a set sample size. This is somewhat different to converging to a set CI width with a variable sample size. Intuitively, it would be expected that the coverage probability would be close to the nominal CL and Ross [25] noted that the coverage probability would match the nominal CL for sufficiently large n.

Table 11 – CI coverage for the convergence statistics for a range of tolerances.

Statistic	Tolerance	Average Sample Size (st.dev)	CI coverage probability
MT	0.06	47.6 (7.6)	94.6% ± 0.4%
	0.04	101.3(15.5)	94.8% ± 0.4%
	0.02	405.5(30.6)	94.8% ± 0.4%
SD	0.4	52.4 (19.0)	91.6% ± 0.5%
	0.2	211.5 (62.2)	93.8% ± 0.5%
	0.1	858.1 (150.3)	93.9% ± 0.5%
ERD	0.04	40.6(2)	97.9% ± 0.3%
	0.02	122.1(17)	94.3% ± 0.5%
	0.01	484.4 (34.4)	94.4% ± 0.5%
EPC	0.08	40.4(1.6)	93.9% ± 0.5%
	0.04	112.8(17.8)	93.5% ± 0.5%
	0.02	456.0 (34.7)	94.2% ± 0.5%
SC	0.02	43.3(3.7)	99.8% ± 0.1%
	0.01	83.6(6.3)	99.5% ± 0.1%
	0.005	161.6(8.8)	98.7% ± 0.2%

As would be anticipated from section 5.1 and 5.2, it can also be seen that as the tolerance decreases the number of samples required for convergence increases. This increase is approximately quadratic for all the measures apart from the SC. For the SC, as the tolerance decreases there is a linear increase in the number of samples required for convergence.

6 Discussion

The advantage of the convergence scheme developed here compared to Ronchi et al.'s scheme is the convergence is directly based on the required precision of the statistics of interest. The most intuitive tolerance to set is Tol_{MT}. It has been previously suggested that the tolerance for Ronchi et al.'s original method could be based on uncertainty from either a fire modelling study determining available safe egress time or uncertainty based on experimental egress trials [19]. Those techniques could also be applied to the CI based method provided those uncertainties are calculated using CIs. In lieu of that type of information, a typical choice for Tol_{MT} could be 0.02, the equivalent of ± 1% error, but the choice depends on the user's requirements. However, the choice of tolerances for the other statistics SD, ERD, EPC and, SC are less apparent.

Prescribing appropriate tolerances for the ERD, EPC and SC FAM-based CIs is problematic, but the case study suggests a potential way forward. It was found that the W_{EPC}, with a 95% confidence level, and W_{MT} have a similar magnitude, W_{EPC} is approximately 10% wider than W_{MT}, for the same sample size (see Table 10) and the behaviour of the corrected CI limits of MT and EPC (see Fig. 7 and Fig. 10) are also similar. It was also shown that under certain conditions the true error measured by EPC and MT are approximately equivalent (see appendix). The ERD also has a behaviour and magnitude related

to MT under certain circumstances. W_{ERD} is roughly half the magnitude of W_{MT} for the same sample size (see Table 10) and the upper confidence limit of the ERD and MT CIs are similar in magnitude and trend (see Fig. 7 and Fig. 10). It was also shown (see appendix) that the true error measured by ERD and MT are also equivalent under certain circumstances. In lieu of any other requirements, it is reasonable to base the tolerances of the ERD and EPC CIs on the specified tolerance of MT (eq. 66 and eq. 67). Although there is no obvious relationship between MT and SC convergence behaviour, it is also suggested that Tol_{SC} is also based on the specified tolerance of MT (eq. 68) unless the user has a specific requirement.

$$Tol_{\text{EPC}} = Tol_{\text{MT}} \tag{66}$$

$$Tol_{\text{ERD}} = \frac{Tol_{\text{MT}}}{2} \tag{67}$$

$$Tol_{\text{SC}} = \frac{Tol_{\text{MT}}}{2} \tag{68}$$

Setting the Tol_{SD} in relation to Tol_{MT} is not recommended. From Table 8, even with 4000 simulations the average W_{SD} is 0.049. By assuming that the TETs follow a normal distribution it can be calculated [22] that it would take ~10000 simulations to achieve convergence for a Tol_{SD} of 0.02. Unless the user specifically requires a certain precision for the SD then it is suggested setting Tol_{SD} between 0.4 and 0.2.

There are circumstances where a user would wish to set the MT tolerance based on time, e.g. ±10s, $Tol_{MT,seconds} = 20s$. The tolerance of the FAM-based CIs cannot be directly expressed in seconds, but this is not a particular problem as the $Tol_{MT,seconds}$ is easily converted to a decimal (eq. 69). This needs to be calculated at runtime as MT is generally unknown a priori.

$$Tol_{\text{MT}} = \frac{Tol_{\text{MT, seconds}}}{MT} \tag{69}$$

In the previous case study, $W_{EPC}/2$ and W_{ERD} are slightly larger than $W_{MT}/2$. As W_{ERD} is greater than $W_{MT}/2$ then there is a greater relative variability expressed in the **AC** compared to the MT (see appendix). This is due to the nature of the theoretical case study model. The coefficient of variation (CV) of the exit time of the kth agent, for the case study model, decreases with increasing k (see eq. 70). Thus, the earlier exiting agents exhibit the most relative variability. However, the actual variability of the kth agent, i.e. SD_k, increases with increasing k. The ERD gives greater weight to larger differences due to the squaring term (eq. 22) leading to a small increase in W_{ERD} compared to $W_{MT}/2$. When performing practical egress simulations, there is potentially larger variation in the exit time of agents other than the last agent to exit which would be reflected in the FAM-based CIs of the **AC**.

$$CV_k = \frac{SD_k}{\bar{t}_k} = \frac{\sqrt{k \times 180}}{k \times 12} = \frac{\sqrt{5}}{2\sqrt{k}} \tag{70}$$

The CI based convergence scheme has a higher computational cost than Ronchi et al.'s original method. This includes generating the bootstrap samples, ordering the values, and calculation of the BCa coefficients. However, it is anticipated that the overall cost will be small compared to the computational cost of the evacuation simulations. If bootstrapping takes a significant amount of time, relative to the evacuation simulations, then the algorithm can be optimised by increasing the number of simulations (K) performed between convergence checks. The calculations have been ordered in the convergence scheme so that the most computationally expensive bootstrapping, the FAM-based CIs, are not calculated until the MT and SD have been deemed to have converged.

The CI method also requires additional memory compared to their original method, although this is not expected to be a serious limitation. The largest memory requirement is the need to store all n egress curves. For example, a 'large' problem consisting of a scenario with 100,000 agents that is repeated

1000 times would require less than 500MB of RAM. This is a modest memory requirement for a modern computer.

Although the use of CIs addresses Ronchi et al.'s first noted limitation, the other limitations they discussed still apply to the work performed here. For instance, it is possible that egress curves can appear identical even when the dynamics of the evacuation are different. Furthermore, differing egress curves compared to the **AC** can return the same ERD, EPC or SC measures. The convergent behaviours of the FAM-based CIs of the **AC** are therefore imperfect proxies for ensuring all the variability in an egress simulation is adequately represented. However, these measures are superior to using the convergence of the MT alone when trying to determine whether enough simulations have been performed to represent the true variability of the evacuation scenario.

7 Concluding Comments

Ronchi et al. [17] pioneered the use of the convergence behaviour of the **AC** to better represent the convergence of the evacuation scenario in general. In this paper, their original approach was modified to use CIs. The CI approach has been shown to give reliable convergence with reference to the estimate interval of the *population* parameters when applied to the mean TET, standard deviation of TETs and the **AC**. The FAM-based CIs for the **AC** have been calculated using a novel application of bootstrapping, FAMs, and bisection algorithm. Although the methods described in this paper are more complex than Ronchi et al.'s original convergence indicators, the resultant CIs have a more standard statistical interpretation than their original convergence indicators. The choice of convergence tolerances is therefore more straightforward as it is just the required statistical precision of the statistics of interest. The case study and algebraic study (see appendix) of the ERD and EPC true errors showed there can be equivalency between those errors and the error in MT. These studies provided guidance for setting the tolerances of ERD and EPC.

The convergence scheme is easily adapted to include other statistics of interest. There are many more parameters that may be of interest, e.g. measures of congestion, that may not have a known CI. For example, Smedburg [45] suggested several statistics, such as egress exit flow rates, which were represented by time series data; FAM-based CIs and convergence for those statistics could be determined using the methods described in this paper. The calculation of the CIs and convergence should be implemented in software so that end users can have easy access to the technology.

The CIs can also be easily applied to experimental or simulated trial data without the convergence scheme suggested here where the number of trials performed is limited by other criteria.

Acknowledgements

Appendix – Equivalency of EPC and ERD errors of the mean egress curve with the error in mean total evacuation time under certain conditions

Under certain conditions it can be shown that the normalised errors, ε_{EPC} and ε_{ERD}, between the sample mean egress curve $\mathbf{AC}^{(n)}$ and the population mean egress curve $\mathbf{AC}^{(\infty)}$, in EPC and ERD are equivalent to the normalised error, ε_{MT}, between the sample mean TET ($MT^{(n)}$) and population mean TET ($MT^{(\infty)}$). The ε_{MT} is given by eq. 71.

$$\varepsilon_{\text{MT}} = \frac{MT^{(\infty)} - MT^{(n)}}{MT^{(n)}} \tag{71}$$

The ε_{EPC} can be expressed as eq. 72.

25

$$\varepsilon_{\text{EPC}} = EPC\big(\mathbf{AC}^{(\infty)}, \mathbf{AC}^{(n)}\big) - 1 = \frac{\left(\sum_{k=1}^{m} ac_k^{(\infty)} ac_k^{(n)}\right)}{\left(\sum_{k=1}^{m} ac_k^{(n)} ac_k^{(n)}\right)} - 1 \tag{72}$$

Substituting $ac_k^{(\infty)}$ with $\gamma_k MT^{(\infty)}$ and $ac_k^{(n)}$ with $\lambda_k MT^{(n)}$ into eq. 72 leads to eq. 73. Note that $ac_m \equiv MT$.

$$\varepsilon_{\text{EPC}} = \frac{\left(\sum_{k=1}^{m} \gamma_k MT^{(\infty)} \lambda_k MT^{(n)}\right)}{\left(\sum_{k=1}^{m} (\lambda_k MT^{(n)})^2\right)} - 1 \tag{73}$$

Simplifying eq. 73 leads to eq. 74.

$$\varepsilon_{\text{EPC}} = \frac{\left(\left(\sum_{k=1}^{m-1} \gamma_k \lambda_k\right) + 1\right) MT^{(\infty)}}{\left(\left(\sum_{k=1}^{m-1} \lambda_k^2\right) + 1\right) MT^{(n)}} - 1 \tag{74}$$

When eq. 75 is satisfied then ε_{EPC} will be approximately equal to ε_{MT} (eq. 76).

$$\left(\left(\sum_{k=1}^{m-1} \gamma_k \lambda_k\right) + 1\right) \approx \left(\left(\sum_{k=1}^{m-1} \lambda_k^2\right) + 1\right) \tag{75}$$

$$\varepsilon_{\text{EPC}} \approx \frac{MT^{(\infty)}}{MT^{(n)}} - 1 = \frac{MT^{(\infty)} - MT^{(n)}}{MT^{(n)}} = \varepsilon_{\text{MT}} \tag{76}$$

There are two specific examples, of when eq. 75 is satisfied, provided here. The first example is when eq. 77 is satisfied. The second example is when both eq. 78 and eq. 79 are satisfied.

$$\gamma_k \approx \lambda_k \ \forall k \tag{77}$$

$$\sum_{k=1}^{m-1} \gamma_k \lambda_k \ll 1 \tag{78}$$

$$\sum_{k=1}^{m-1} \lambda_k^2 \ll 1 \tag{79}$$

The ε_{ERD} (eq. 80) can be expressed as eq. 81.

$$\varepsilon_{\text{ERD}} = ERD\big(\mathbf{AC}^{(\infty)}, \mathbf{AC}^{(n)}\big) = \sqrt{\frac{\sum_{k=1}^{m} \left(ac_k^{(\infty)} - ac_k^{(n)}\right)^2}{\sum_{k=1}^{m} \left(ac_k^{(n)}\right)^2}} \tag{80}$$

$$\varepsilon_{\text{ERD}} = \sqrt{\frac{\left(\sum_{k=1}^{m-1} \left(ac_k^{(\infty)} - ac_k^{(n)}\right)^2\right) + \left(MT^{(\infty)} - MT^{(n)}\right)^2}{\left(\sum_{k=1}^{m-1} \left(ac_k^{(n)}\right)^2\right) + \left(MT^{(n)}\right)^2}} \tag{81}$$

Two examples of approximate equivalency of ε_{ERD} and ε_{MT} (eq. 82) are given here. The first example occurs when eq. 83 is true (note: eq. 84). The simplest example of eq. 83 being satisfied is when eq. 85 is true. The second example is when both eq. 86 and eq. 87 are true.

$$\varepsilon_{\text{ERD}} \approx \left|\frac{MT^{(\infty)} - MT^{(n)}}{MT^{(n)}}\right| = |\varepsilon_{\text{MT}}| \tag{82}$$

$$\frac{\sum_{k=1}^{m-1} \left(ac_k^{(\infty)} - ac_k^{(n)}\right)^2}{\sum_{k=1}^{m-1} \left(ac_k^{(n)}\right)^2} \approx \frac{\left(MT^{(\infty)} - MT^{(n)}\right)^2}{\left(MT^{(n)}\right)^2} \tag{83}$$

$$\frac{x}{y} \equiv \frac{\epsilon x}{\epsilon y} \equiv \frac{x + \epsilon x}{y + \epsilon y} = \frac{x(1 + \epsilon)}{y(1 + \epsilon)} \tag{84}$$

$$\frac{\left|ac_k^{(\infty)} - ac_k^{(n)}\right|}{ac_k^{(n)}} \approx \frac{\left|MT^{(\infty)} - MT^{(n)}\right|}{MT^{(n)}} \forall k \tag{85}$$

$$\left(\sum_{k=1}^{m-1}\left(ac_k^{(\infty)} - ac_k^{(n)}\right)^2\right) \ll \left(MT^{(\infty)} - MT^{(n)}\right)^2 \tag{86}$$

$$\left(\sum_{k=1}^{m-1}\left(ac_k^{(n)}\right)^2\right) \ll \left(MT^{(n)}\right)^2 \tag{87}$$

Figure 12 depicts two sets of contrived egress curves with their errors reported in Table 12. The first set of curves satisfy both eq. 77 and eq. 85 and all the errors are equal (0.2). The second set of curves satisfy eq. 78 and eq. 79 and, eq. 86 and eq. 87 and ε_{ERD} and ε_{EPC} are approximately equal to ε_{MT} (0.19 $\approx$ 0.2).

Table 12 – MT, EPC and ERD Error for two sets of contrived curves (Fig. 11). The values demonstrate the equivalency of the error measures under certain conditions.

Set of curves	ε_{MT}	ε_{ERD}	ε_{EPC}
1	0.2	0.2	0.2
2	0.2	0.19	0.19
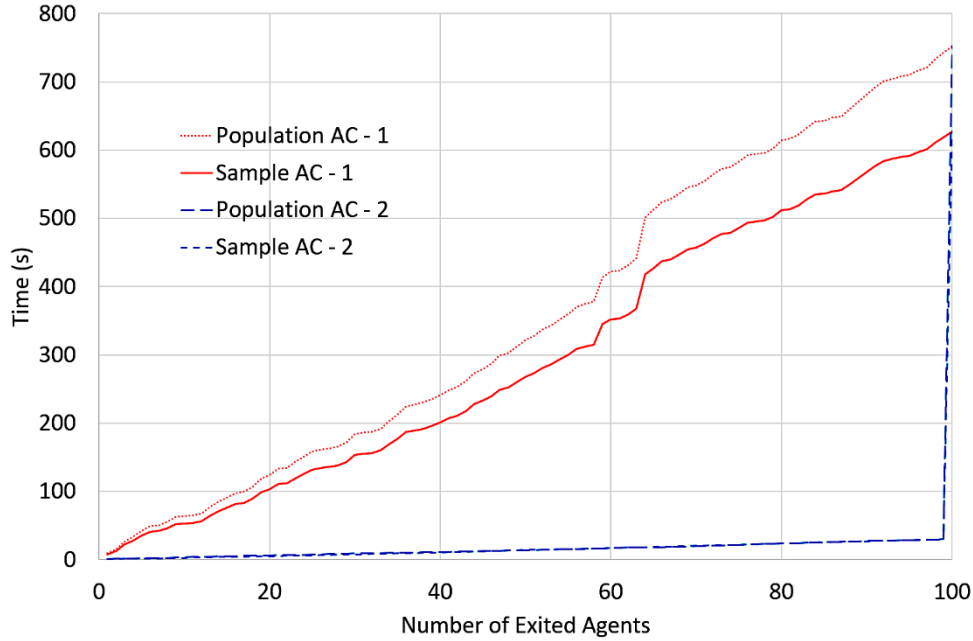
Fig. 12 – Two sets of contrived curves that demonstrate the approximate equivalency of ε_{ERD}, ε_{EPC} and ε_{MT}. $MT^{(n)}$ is 627s for both sample $\mathbf{AC}^{(n)}$ curves and $MT^{(\infty)}$ is 752s for both population $\mathbf{AC}^{(\infty)}$ curves.

By extending the analysis of the ERD relationship it can also be seen, from eq. 81, that if eq. 88 is true then ε_{ERD} will be greater than $|\varepsilon_{MT}|$, indicating there is more relative difference in the rest of the egress curve compared to the TET. One specific circumstance of this occurs when eq. 89 is true. Similarly, if eq. 90 is true then ε_{ERD} will be less than $|\varepsilon_{MT}|$, showing there is less relative difference in the rest of the egress curve compared to the TET. One specific circumstance of this occurs when eq. 91 is true.

$$\frac{\sum_{k=1}^{m-1}\left(ac_k^{(\infty)} - ac_k^{(n)}\right)^2}{\sum_{k=1}^{m-1}\left(ac_k^{(n)}\right)^2} > \frac{\left(MT^{(\infty)} - MT^{(n)}\right)^2}{\left(MT^{(n)}\right)^2} \tag{88}$$

$$\frac{\left|ac_k^{(\infty)} - ac_k^{(n)}\right|}{ac_k^{(n)}} > \frac{\left|MT^{(\infty)} - MT^{(n)}\right|}{MT^{(n)}} \forall k \tag{89}$$

$$\frac{\sum_{k=1}^{m-1}\left(ac_k^{(\infty)} - ac_k^{(n)}\right)^2}{\sum_{k=1}^{m-1}\left(ac_k^{(n)}\right)^2} < \frac{\left(MT^{(\infty)} - MT^{(n)}\right)^2}{\left(MT^{(n)}\right)^2} \tag{90}$$

$$\frac{\left|ac_k^{(\infty)} - ac_k^{(n)}\right|}{ac_k^{(n)}} < \frac{\left|MT^{(\infty)} - MT^{(n)}\right|}{MT^{(n)}} \forall k \tag{91}$$

References

[1]   Gwynne, S., Galea, E.R., Owen, M., Lawrence P.J., and Filippidis L. (1999) A Review of the Methodologies Used in Evacuation Modelling, Fire and Materials 23(6): 383-389, http://dx.doi.org/10.1016/s0360-1323(98)00057-2

[2]   Kuligowski, E.D., Peacock, R.D., and Hoskins, B.L. (2010) A Review of Building Evacuation Models, 2nd Edition, National Institute of Standards and Technology (NIST) Technical Note 1680

[3]   Gwynne, S., Galea, E.R., Lawrence, P.J., and Filippidis, L. (2001) Modelling Occupant interaction with fire conditions using the building EXODUS evacuation model, Fire Safety Journal 36: 327-357, http://dx.doi.org/10.1016/S0379-7112(00)00060-6

[4]   Ha S., Ku N.K., Roh M.I., and Lee K.Y. (2012) Cell-based evacuation simulation considering human behavior in a passenger ship. Ocean Eng 53: 138-152. http://doi.org/10.1016/j.oceaneng.2012.05.019

[5]   Korhonen, T., Hostikka, S., Heliovaara, S., and Ethamo, H. (2008) FDS+Evac: An agent based fire evacuation model, Pedestrian and Evacuation Dynamics, Springer, Wuppertal, Germany, pp 109–120. http://dx.doi.org/10.1007/978-3-642-04504-2_8

[6]   Meyer-König, T., Valanto, P., and Povel, D. (2005) Implementing ship motion in AENEAS — model development and first results. In Pedestrian and Evacuation Dynamics, ed. N. Waldau, P. Gattermann, H. Knoflacher, and M. Schreckenberg, pp 429-441, http://dx.doi.org/10.1007/978-3-540-47064-9_41

[7]   Park, J.H., Lee, D.K., Kim, H.T., Yang, Y.S. (2004) Development of evacuation model for human safety in maritime casualty, Ocean engineering. 31(11): 1537-1547, http://dx.doi.org/10.1016/j.oceaneng.2003.12.011

[8]   Thompson, P.A., and Marchant, E.W. (1995) A computer model for the evacuation of large building populations, Fire Safety Journal 24(2): 131-148, ISSN 0379-7112, http://dx.doi.org/10.1016/0379-7112(95)00019-P

[9]   Averill, J.D. (2011) Five grand challenges in pedestrian and evacuation dynamics, Proceedings of the 5th international pedestrian and evacuation dynamics conference, pp 1-11, http://dx.doi.org/10.1007/978-1-4419-9725-8_1

[10]   Lovreglio R., Spearpoint M., and Girault M. (2019) The Impact of Sampling Methods on Evacuation Model Convergence and Egress Time, Reliability Engineering and System Safety, DOI: 10.1016/j.ress.2018.12.015

[11]   Li, L., Wei, L. L., Kinsey, M., Sun, A. (2018) HSTEAM – High Speed Train Evacuation Analysis Methodology, Procedia Engineering (2018) 18-27 https://doi.org/10.1016/j.proeng.2017.12.133

[12]   Revised Guidelines for Evacuation Analysis for New and Existing Passenger Ships, IMO MSC.1/Circ 1533, 6 June 2016.

[13]   Grandison, A., Deere, S., Lawrence, P., and Galea, E.R. (2017) The use of confidence intervals to determine convergence of the total evacuation time for stochastic evacuation models, Ocean Engineering 146: 234–245, https://doi.org/10.1016/j.oceaneng.2017.09.047

[14]   Galea, E.R. (2006) Proposed Methodology for the Use of Computer Simulation to Enhance Aircraft Evacuation Certification, AIAA Journal of Aircraft, Vol 43, Number 5, pp 1405-1413, 2006. DOI: 10.2514/1.20937

[15]   Meacham, B., Lord, J., Moore, A., Fahy, R., Proulx, G., Notarianni, K. (2004) Investigation of Uncertainty in Egress Models and Data, Proceedings of the 3rd International Symposium on Human Behaviour in Fire, Belfast, UK, September 01-03, 2004, pp. 419-428, (NRCC-47308)

[16]   Ronchi, E. (2014) The need for a verification and validation protocol for evacuation models, Fire and Evacuation model. tech. conf., Gaithersburg MD (USA), DOI: 10.13140/2.1.4893.3127

[17]   Ronchi, E., Reneke, P.A., Peacock, R.D. (2014) A method for the analysis of behavioural uncertainty in evacuation modelling, Fire Technology 50: 1545-1571

[18]   Ronchi, E. & Nilsson, D. (2014) Modelling total evacuation strategies for high-rise buildings. Build. Simul. doi:10.1007/s12273-013-0132-9

[19]   Lovreglio R., Ronchi E., and Borri D. (2014) The validation of evacuation simulation models through the analysis of behavioural uncertainty, Reliability Engineering & System Safety 131: 166-174, http://dx.doi.org/10.1016/j.ress.2014.07.007

[20]   Jullien, Q., Paillat, J-L., Thiry-Muller, A., Lardet, P., Pinoteau, N. (2019) Use of statistical approach on stochastic building egress simulations applied to buildingEXODUS and Pedestrian Dynamics, proceedings of Interflam 2019, pp705-715.

[21]   Neyman, J. (1937) Outline of a theory of statistical estimation based on the classical theory of probability, Philosophical Transactions of the Royal Society of London, Series A 236: 333–380, doi:10.1098/rsta.1937.0005

[22]   Bluman, A. G. (2007) Elementary Statistics: A step by step approach (6th Edition), New York, McGraw-Hill, ISBN 978-0-07-304825-3.

[23]   Byrne, M.D. (2013) How many times should a stochastic model be run? An approach based on confidence intervals, Proceedings of the 12th international conference on cognitive modeling, Ottawa, pp 445-450

[24]   Driels, M.R., and Shin, Y.S. (2004) Determining the number of iterations for Monte Carlo simulations of weapon effectiveness, Technical Report NPS-MAE-04-005, Naval Postgraduate School, Defense Threat Reduction Agency, April 2004

[25]   Ross, S.M. (2013) Simulation. Academic Press, ISBN 9780124158252. https://doi.org/10.1016/C2011-0-04574-X

[26]   Winston, W.L. (2000) Simulation Modeling using @RISK, Duxbury Press, ISBN 978-0534380595.

[27]   Jackel, P. (2003) Monte Carlo methods in finance, John Wiley & Sons Inc, ISBN 978-0471497417.

[28]   Oberle, W. (2015) Monte Carlo Simulations: Number of Iterations and Accuracy, ARL-TN-0684, US Army Research Laboratory, July 2015.

[29]   Galea E.R., Deere S., Brown R., Filippidis L. (2014) An Evacuation Validation Data Set for Large Passenger Ships. In: Weidmann U., Kirsch U., Schreckenberg M. (eds) Pedestrian and Evacuation Dynamics 2012. Springer, Cham

[30]   Peacock, R.D., Reneke, P.A, Davis, W.D, and Jones, W.W. (1999) Quantifying fire model evaluation using functional analysis, Fire Safety Journal 33(3): 167-184

[31]   Barrett, J.P., and Goldsmith, L. (1976) When is n sufficiently large?, American Statistician 30(2): 67-70

[32]   Efron, B., and Tibshirani, R. (1993) An Introduction to the Bootstrap. Boca Raton, FL: Chapman & Hall/CRC., ISBN 0-412-04231-2

[33]   Efron, B. (1987) Better bootstrap confidence intervals, Journal of the American Statistical Association, 82(397): 171-185, doi:10.2307/2289144

[34]   Efron, B. (1979) Bootstrap methods: Another look at the jackknife, Annals of Statistics. 7(1): 1–26, doi:10.1214/aos/1176344552

[35]   Davison, A.C., and Hinkley, D.V. (1997) Bootstrap Methods and Their Application. Cambridge, U.K.: Cambridge University Press, doi:10.1080/00031305.2015.1089789

[36]   Hall, P. (1988) Theoretical comparison of bootstrap confidence intervals, Ann. Statist. 16(3): 927-953, doi:10.1214/aos/1176350933.

[37]   Romano, J.P. (1988) Bootstrapping the mode, Ann Inst Stat Math (1988) 40: 565. https://doi.org/10.1007/BF00053066

[38]   Athreya, K.B. (1987) Bootstrap of the mean in the infinite variance case Ann Stats vol 15 (2) 724–731 doi:10.1214/aos/1176350371

[39]   Cavaliere, G., Georgiev, I., and Taylor, A. M. R. (2013) Wild Bootstrap of the Sample Mean in the Infinite Variance Case, Econometric Reviews, 32:2, 204-219, DOI: 10.1080/07474938.2012.690660

[40]   Hesterberg, T.C., What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum, American Statistician 69(4): 371-386, doi: 10.1080/00031305.2015.1089789

[41]   Carpenter, J., and Bithell, J. (2000) Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. Statist. Med. 19: 1141-1164, doi:10.1002/(SICI)1097-0258(20000515)19:9<1141::AID-SIM479>3.0.CO;2-F

[42]   Steele, J.M. (2004) The Cauchy–Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities. The Mathematical Association of America, ISBN 978-0521546775.

[43]   Dunn, O. J. (1961) Multiple Comparisons Among Means, Journal of the American Statistical Association. 56 (293): 52–64. doi:10.1080/01621459.1961.10482090.

[44]   Matsumoto, M., and Nishimura, T. (1998) Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator, ACM Transactions on Modeling and Computer Simulation 8(1): 3–30, doi:10.1145/272991.272995

[45]   Smedberg, E. (2019) The Analysis of Results of Stochastic Evacuation Models, Report 5587 ISRN: LUTVDG/TVBB--5587—SE, Lund University, Sweden