REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED** START DATE	END DATE
MARCH 2024	TECHNICAL PAPER	JUNE 2023	DECEMBER 2023

4. TITLE AND SUBTITLE
TESTING CONFIDENCE INTERVALS ON THE DURATION OF AN EVENT BASED ON THE COPERNICAN PRINCIPLE

5a. CONTRACT NUMBER	**5b. GRANT NUMBER**	**5c. PROGRAM ELEMENT NUMBER**
IN-HOUSE	N/A	62788F
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**
RIGIHRDR		R1LP

6. AUTHOR(S)
Warren H. Debany Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory/Information Directorate
Rome Research Site/RIG
525 Brooks Road
Rome NY 13441-4505

8.PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory/Information Directorate
Rome Research Site/RIG
525 Brooks Road
Rome NY 13441-4505

10.SPONSOR/MONITOR'S ACRONYM(S)
AFRL/RI

11.SPONSOR/MONITOR'SREPORT NUMBER(S)
AFRL-RI-RS-TP-2024-010

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited. PA# AFRL-2024-0210
Date Cleared: 12 Jan 2024

13. SUPPLEMENTARY NOTES

14. ABSTRACT
J. R. Gott proposed a method for placing confidence intervals (CIs) on the remaining duration of an ongoing event, where an event can be a phenomenon, process, failure-free operation of an item, or existence of an object. Gott's method can provide a lower bound and upper bound on event duration without any knowledge of its underlying distribution: it requires knowing only the past duration of the event. Its only necessary assumption is that the “Copernican Principle” holds, meaning that the observation of the ongoing event occurs at “no special time.” The bounds on the remaining duration of an event obtained by Gott's method are very wide, being 1/39 and 39 times the observed past duration for a 95% CI; other techniques that use knowledge of the underlying distribution can yield much tighter bounds, but such information may not be known or even knowable. Gott's method was tested using Monte Carlo simulation with populations of events that satisfied the Copernican Principle and it was shown to yield 95% CIs regardless of the underlying distribution of event durations. It was also tested using “baby boom” populations that violated that principle and it was found to be tolerant of those violations. Gott's 95% CIs can also be continuously recalculated throughout the entire duration of an event (another violation of the Copernican Principle) and the true end time will still be found to have been within those bounds for 95% of that duration. The lower bound provided by Gott's method may closely approximate that obtained from the conditional survival method that uses the true underlying distribution; this is significant because the lower bound on remaining duration can be a useful measure of the statistical level of trust that a critical function will not fail within a quantifiable period of time.

15. SUBJECT TERMS
Confidence Intervals, Survival Probability, Conditional Survival, Reliability, Trust, Trustworthiness

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	C. THIS PAGE	**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**
U	U	U	SAR	90

19a. NAME OF RESPONSIBLE PERSON	**19b. PHONE NUMBER** *(Include area code)*
JAMES S. PERRETTA	**N/A**

PREVIOUS EDITION IS OBSOLETE.
STANDARD FORM 298 (REV. 5/2020)
Prescribed by ANSI Std. Z39.18

Testing Confidence Intervals On the Duration of an Event Based on the Copernican Principle

Warren H. Debany Jr., PhD
Volunteer Emeritus Corps (VEC)
Information Warfare Division (AFRL/RIG)
Air Force Research Laboratory
525 Brooks Rd.
Rome NY 13441-4505

Abstract

J. R. Gott proposed a method for finding a confidence interval for the remaining duration of an event that is observed to be ongoing at a certain time. An event can be a phenomenon, process, operation of an item, or existence of an object, which has a duration with a start and an end.

Gott's method requires knowing only the past duration of the event when observed, and its only necessary assumption is that the "Copernican Principle" holds. In this context, this principle means that the observation of an ongoing event occurs at no special time in the duration of the event.

The bounds on the remaining duration obtained by Gott's method are very wide, being 1/39 and 39 times the observed past duration for a 95% confidence interval. Techniques that use additional information or even the true underlying distribution can yield much tighter bounds, but such information may not be known and in the case of a one-time or one-of-a-kind event may be unknowable.

This paper reports the results of testing Gott's method using Monte Carlo simulation with populations of events that satisfy the Copernican Principle. Experiments showed that Gott's method indeed provides 95% confidence intervals regardless of the underlying distribution of event durations. The confidence intervals have equal tail probabilities so they are central as well.

Gott's method was also shown to have some tolerance of violations of the Copernican Principle, specifically a "baby boom" model for event start times. The actual or measured confidence level of the confidence intervals in these experiments tended to stay close to the target level of 95%, although the confidence intervals were weaker as they were slightly non-central with unknown tail probabilities.

It was also demonstrated that Gott's confidence intervals can be continuously recalculated throughout the entire duration of an event and, retrospectively, the true duration will be found to have been within those recalculated bounds for 95% of that time.

It was noted that the lower bound provided by Gott's method may often closely approximate the lower bound obtained from the conditional survival method that uses the true underlying distribution. This is significant because the lower bound on remaining duration can be a useful measure of the statistical level of trust that a critical function will not fail within a quantifiable period of time.

Contents

1 Introduction

The information contained in this publication is for research purposes only and is not intended as a form of direction or advice, or application to any specific use case.

The views presented are those of the author and do not necessarily represent the views of DoD or its components.

The use of the name or mark of any specific manufacturer, commercial product, commodity, or service in this publication does not imply endorsement by DoD or its components.

Albert Einstein is frequently credited with saying, “The definition of insanity is doing the same thing over and over and expecting different results.” It seems there is no evidence that Einstein ever said it,[1] but even so, this advice is used to justify making only one attempt at efforts in fields such as science and engineering, business, and organizational behavior. Whether a result was success or failure, invoking this statement implies that there is no variation in life or nature and future outcomes will always be the same as the first.

No statement can be further from the truth. Probability and statistics developed as fields of mathematics driven by the fact that variation is the norm and not the exception. In most interesting real-world cases, when samples from a population of items or outputs of a process are observed, results tend to vary from item to item or observation to observation.

Confidence Intervals (*CIs*) are a fundamental statistical tool for understanding the scope or limits of variability. Based on one or more observations, one can place bounds on estimates of the underlying parameters of a process or population. Similarly, when the parameters are known, CIs can place bounds on the values of future observations.

[1] See discussions at https://www.history.com/news/here-are-6-things-albert-einstein-never-said, and https://quoteinvestigator.com/2017/03/23/same/.

This author recently completed a study [Debany (2023)] of CIs for the unknown probability of success[2] in an infinite population or process, or the number of successes in a finite population, based on multiple samples from that population. Rare events were emphasized, where the probability or number of successes was small. Sampling With Replacement (SWR) and Sampling Without Replacement (SWoR) were considered.

The main focus of the study was to determine the quality of *exact confidence intervals* compared to the quality of published approximations, where exact CIs are obtained by summing values of the probability mass functions of the governing discrete distributions. Binomial, negative binomial, hypergeometric, and negative hypergeometric distributions were used.

With k successes observed in a sample of size n, Lower Bounds (LBs) and Upper Bounds (UBs) were obtained for the probability of success p for SWR and for the number of successes r in a population of size N for SWoR. For the most part, two-sided 95% central CIs[3] were obtained.

This current paper addresses a problem that is complementary to the aforementioned study. Instead of finding CIs for the success probability or number of successes in a population based on multiple samples, this paper tests a method for obtaining CIs for the "successful" or remaining duration of an *ongoing event* based on a single observation or sample.

The problem and solution tested here were introduced by J. R. Gott in his 1993 paper "Implications of the Copernican Principle for Our Future Prospects" [Gott (1993)].

[2] A "success" may be a desirable or undesirable outcome. For example, if the concern is the prevalence of defective items in a population, one "successfully" finds a bad item.

[3] A *two-sided* CI has a lower and an upper bound, and with those values is denoted as [LB, UB].

A two-sided 95% CI means that the parameter or sampled value of interest has a 95% probability of being within the LB and UB. The probability of being below the LB is the *lower tail probability*, and the probability of being above the UB is the *upper tail probability*. The sum of the two tail probabilities is the *level of significance* which in this case is 5%.

Two-sided CIs are called *central* when their lower and upper tail probabilities are equal, so a central 95% CI would have tail probabilities of 2.5%. Two-sided CIs are called *non-central* when the two tail probabilities are not equal. For example, see [Pires (2008)].

The concept of an "event" is treated abstractly in this paper. An event of interest could be defined by a natural phenomenon that appears and disappears. It could represent the lifetime of a piece of equipment put into operation to perform a critical function and eventually fails. It could be the duration of a structure that is built and eventually decays or is destroyed.

An event has a total *duration*, also called the *lifetime* or *time to failure* in this paper. An event duration has a known *start time* and it has an eventual but unknown *end time* that is the point at which the event will stop, fail, be destroyed, disappear, or otherwise terminate.

An event must have a duration greater than 0, or it could not be observed, and it is reasonable to believe that the duration of an event with a known and thus finite start time must be finite as well.

Determining the end time of an event may be equivocal, however, when the ending is not sharp or discrete. The degrading condition of a phenomenon or an item may require the setting of a threshold or other criterion that defines the soft ending of an event.[4] However, even events that would end only gradually or imperceptibly under normal circumstances may come to an abrupt finish; this is discussed in Section 4.

In the approach used here, an event is observed to be ongoing when an *observer* takes a *single sample* from a population or occurrence of a process. There may be many events in the population, or the entire population might consist of a one-time event.

The event is captured by the observer at the *sample time*. The *observation time* or *past duration*, denoted by t_{obs}, is the difference between the start time and the observer's sample time. The total duration of the event from the start time until the end time is denoted by t_{end}. The *remaining duration* or *future duration*, denoted by t_{rem}, is the difference between the observation time and the end time.[5]

[4] Events may have soft beginnings as well, which would also require subjective definitions.

[5] To "sample," "observe," or "capture" an event are synonymous in this paper, with the convention that start time and sample time (as well as the unknown end time) are measured in calendar time (or "simulation calendar time" for these experiments), while the observation time and duration are measured from the start of the event.

The observer's[6] goal is to find the LB and UB of a CI for the remaining duration of the event or, equivalently, the total duration. Once an event has been captured, only the observation time t_{obs} is needed for the purpose of obtaining a CI for t_{rem} or t_{end}. The actual start time and sample time can then be disregarded.

This paper tests two-sided CIs with the often-used *confidence level* of $C = 0.950$ or 95%,[7] that being the probability of the duration or remaining duration being between the LB and UB of the CI. The confidence level of 95%, meaning a 5% level of significance, is used throughout this paper because it is a reasonable value for many applications and so that this paper's results can be compared to related work.[8]

A CI is not a guarantee, as a 95% CI is *intended* to have about 5% of its outcomes outside that interval.

When the duration of an event describes a desirable condition, such as successful operation of an item, bounds on duration can be used as measures of trustworthiness. Both bounds can be useful statistical measures of trustworthiness, with the LB surely being the more significant of the two in this context.

There are many definitions of trust and trustworthiness. A definition that is actionable and consistent with common use is given by Varshney [Varshney (2022)] as:

> "...the idea that the trustee has certain properties that make it *trustworthy*, i.e., the qualities by which the trustor can expect the trustee to perform the important action referred to in the definition of trust."

[6] For simplicity, assume that the observer is also the agent who calculates the CIs based on the observation of the ongoing event.

[7] Decimal fractions and percentages are used interchangeably in the most convenient format.

[8] As noted in this author's previous paper [Debany (2023)], the 5% level of significance was suggested by R. A. Fisher, among others, and although it is used almost universally it has also been the subject of debate and disagreement. Cowles and Davis [Cowles (1982)] gave a history of the 5% level of significance that indicated that this value may be arbitrary but it is a satisfactory threshold. An example of the thought process is that a 10% level of significance implies that a result is not very improbable, and a 1% level of significance implies that a result is very improbable, and the value of 5% is the roughly the midpoint of these two thresholds.

If a given item is the "trustee" and the event of its failure-free operation is the "important action," then the LB on the duration of this event quantifies a length of time during which the observer or "trustor" can expect this to be true with high confidence. The UB, on the other hand, quantifies the outside limit of trustworthiness because failure is likely to occur by that time.

The method tested in this paper was proposed by J. R. Gott [Gott (1993)] for finding such CIs based on a single observation of the current duration of an ongoing event. The significance of this method is that it can be applied when the underlying distribution of the event durations or other prior information is unknown to the observer; one can say furthermore that it should be applied only when this is the case.

No new mathematical principles are presented in this paper. The contribution offered here is solely the empirical testing of Gott's method and the documenting of those results.

The rest of the paper is outlined as follows.

Section 2 outlines Gott's method for obtaining CIs for the remaining duration of an ongoing event based solely on the observation of its past duration.

Section 3 describes the conditional survival method for CIs based on effectively perfect knowledge of the underlying distribution of event durations. This method is the baseline for comparison to Gott's method.

Section 4 discusses an exception to the presumption that the observer has no knowledge of the underlying distribution of event durations for Gott's method. This is where it is reasonable to believe that lifetime is determined more by active operation than by dormant storage or standby conditions.

Section 5 outlines the Monte Carlo simulation approach used to test the CIs.

Section 6 applies the Monte Carlo simulation approach to constant populations of events with steady state or uniformly distributed start times that satisfy the assumption that the observed event is sampled at no special time in the duration of the event.

Section 7 applies the Monte Carlo simulation approach to populations of events that violate the Copernican Principle to some extent with "baby booms" of non-uniformly distributed start times.

Section 8 considers the case where the Copernican Principle is further abandoned and the observer continuously recalculates the CIs throughout the event duration.

Section 9 summarizes the results and conclusions.

2 Confidence Intervals Based on Gott's Method

Gott's method [Gott (1993)] is based on a presumption and a critical assumption:

- Presumed: The true underlying distribution or other prior information pertaining to the duration of the event is unknown to the observer.
- Assumed: The Copernican Principle holds, meaning that the observation occurs at no special time in the duration of the sampled event.

The presumption indicates only when the method should be used, but the assumption is the basis for it. It is shown in this paper, however, that the assumption that the Copernican Principle holds can be violated to a fairly large extent and Gott's method still can give useful results.

The *reliability function*, denoted as $R(t)$, is the probability that the event *has not ended* by time t When a finite event duration is taken to start at $t = 0$, its $R(t)$ has the value 1 at $t = 0$ and falls to 0 as t increases.

The reliability function (or its equivalent) is referred to here as the *underlying distribution* for event duration.[9]

The presumption that the underlying distribution for event duration is unknown to the observer means that this agent does not have or use any knowledge of the event beyond the observation that it is ongoing at the sample time, whether the population consists of many such events or just one. The observer does not use information such as the history, physics, or chemistry of the subject of the event to estimate remaining duration. The observer has no external knowledge that, say, the subject of an event is suddenly in a precarious situation or bulldozers are on their way to demolish it.

[9] Because it is the probability of the event surviving to time t, the reliability function $R(t)$ is also called the *survivor function* or *survival function*.

The complement of the reliability function, $F(t) = 1 - R(t)$, is the *cumulative distribution function (CDF)* for the duration. $F(t)$ is the probability that the event *has ended* by time t and thus is also called the *failure function*.

If it exists, the derivative of $F(t)$, denoted by $f(t)$, is the *probability density function (PDF)* for the duration. If $f(t)$ is discrete, it is called the *probability mass function (PMF)*.

$R(t)$, $F(t)$, and $f(t)$ are equivalent representations of the underlying distribution.

The presumption that $R(t)$ or other prior information is unknown to the observer is not necessary for Gott's method to work; rather, it justifies using this method because no other can be employed. It is demonstrated in this paper that, as would be expected, additional or even full knowledge of the reliability function allows the observer to use techniques such as the conditional survival method, outlined in Section 3, to give far tighter bounds on the duration of an observed event.

The assumption that the *Copernican Principle* holds is the basis of Gott's method, however. Broadly, this is a cosmological assumption that we occupy no special or privileged time or place in the universe [Bondi (1948)] [Gale (1981)] [Rudnicki (1989)], although opinions vary as to the applicability of this principle of homogeneity in terms of uniformity or irregularities of the universe. Regardless of the origins of the term, however, in this context Gott used the "Copernican Principle" in a restricted sense to mean only that the observer samples an ongoing event at no special time in the duration of the event.

This paper uses this term in the same sense.

The Copernican Principle is applied here to mean the assumption that, if an event is observed, then the observation time within that event t_{obs} is distributed as U$[0,\, t_{end}]$, or $\frac{t_{obs}}{t_{end}}$ is distributed as U$[0,\, 1]$, where "U$[a,\, b]$" denotes the uniform distribution on the closed interval $[a,\, b]$.

In other words, t_{obs} is assumed to be uniformly distributed in time from the start to the end of the event. While this is a very stringent assumption, it is the only necessary assumption in Gott's method.[10] And, if the presumption holds that no information about the underlying distribution of event duration is available to the observer, then there may be *no justification to assume any other property* for the distribution of t_{obs} in the event duration.

[10] To be clear, there is also a tacit assumption that no additional agent is involved in this process by manipulating the observer's behavior or sampling opportunities, or by affecting the event duration in reaction to the observer's calculated CI. Such intervention could be beneficial or malicious with respect to the observer but, in any case, it would surely invalidate any analysis that does not explicitly model the interactions involved in that situation.

Two models of observations that satisfy the Copernican Principle are addressed in Section 6.

The first model is the one outlined in the Introduction and is the primary one used in this investigation. This is where a population consists of events that have randomly occurring start times uniformly distributed over some finite timespan, and a single event is sampled from that population by the observer. In many, if not most, cases where Gott's method would be applied, the entire population might consist of a one-time event. In this paper, this is called the *Copernican Sample model*.

The second model is where the observation of an event occurs at a randomly selected point in time that is uniformly distributed over the event duration. In this paper, this is called the *Copernican Observation model*.

Implementation of the Copernican Observation model in the real world outside of simulation would be problematic as it would seem to require some non-causal ability to explicitly observe the ongoing event at a randomly selected time uniformly distributed within a duration. The Copernican Sample model would therefore appear to be the more reasonable of the two, so that is the one emphasized in this paper, but results of simulations based on the Copernican Observation model are presented and discussed briefly as well.

Being based solely on the Copernican Principle, Gott's method yields CIs that have greater length[11] than those that can be obtained using additional information; this is demonstrated using known reliability functions with the conditional survival method outlined in Section 3. This paper shows, however, that Gott's method yields 95% central CIs regardless of the underlying reliability function when the Copernican Principle holds and t_{obs} is uniformly distributed in the event duration.

An unanticipated finding from these Monte Carlo simulations is that the distribution of t_{obs} can deviate substantially from the uniform distribution and Gott's method can still yield close-to-95% CIs. In these cases, unfortunately, the CIs will be non-central.

[11] The *length* (also called the *width*) of a CI is the difference between its LB and UB. For example, see [Brown (2001)] [Pires (2008)]

Gott derived CIs for a two-sided 95% confidence level, but the method works for any desired confidence level and for one-sided CIs as well. His formulation of the method to obtain the bounds for a two-sided CI for the remaining duration of an event is outlined as follows.

For a known past duration t_{obs} and unknown duration t_{end}, let ρ be the ratio of the remaining duration t_{rem} to the past duration t_{obs}, or

$$\rho = \frac{t_{rem}}{t_{obs}} \quad . \tag{1}$$

If the observation time t_{obs} is uniformly distributed in the interval $[0,\ t_{end}]$, then there is a 95% probability that t_{obs} occurs between $0.025\, t_{end} = \frac{1}{40}\, t_{end}$ and $0.975\, t_{end} = \frac{39}{40}\, t_{end}$.

Thus, the LB and UB on ρ for a two-sided 95% CI are given by

$$\rho_{LB} = \frac{t_{end} - \frac{39}{40}\, t_{end}}{\frac{39}{40}\, t_{end}} = \frac{1}{39} \approx 0.0256 \tag{2}$$

$$\rho_{UB} = \frac{t_{end} - \frac{1}{40}\, t_{end}}{\frac{1}{40}\, t_{end}} = 39 \quad . \tag{3}$$

Combining (2) and (3), the two-sided 95% CI for t_{rem}, the remaining duration of the event beyond the observation time t_{obs}, is given by

$$[\rho_{LB}\, t_{obs},\ \rho_{UB}\, t_{obs}] \quad \text{or} \quad [\frac{1}{39}\, t_{obs},\ 39\, t_{obs}] \quad \text{or} \quad [0.0256\, t_{obs},\ 39\, t_{obs}] \quad . \tag{4}$$

Equivalently, Gott's two-sided 95% CI for t_{end}, the total duration of the event, is given by

$$[(1+\rho_{LB})\, t_{obs},\ (1+\rho_{UB})\, t_{obs}] \quad \text{or} \quad [1.0256\, t_{obs},\ 40\, t_{obs}] \quad . \tag{5}$$

Gott's expression (4) for t_{rem}, or its equivalent (5) for t_{end}, is the CI tested in this paper.

This paper tests only two-sided 95% CIs with bounds given by (2) and (3). However, extending this argument for a two-sided CI with confidence level C, Gott's LB and UB on ρ are

$$\rho_{LB} = \frac{1-C}{1+C} \tag{6}$$

$$\rho_{UB} = \frac{1+C}{1-C} \quad . \tag{7}$$

For a *one-sided* CI with confidence level C, Gott's LB and UB on ρ are

$$\rho_{LB} = \frac{1-C}{C} \tag{8}$$

$$\rho_{UB} = \frac{C}{1-C} \quad . \tag{9}$$

For a one-sided 95% CI, (8) and (9) yield $\rho_{LB} = 1/19$ and $\rho_{UB} = 19$.

When the observer has no prior information about the event duration, the only option may be to assume that the Copernican Principle holds. In the case of a one-time or one-of-a-kind phenomenon, such as Gott's example of the duration of the famous and ancient structure Stonehenge [Gott (1993)], the reliability function may not be merely unknown but in fact unknowable.

But even extensive historical data may not be sufficient to provide accurate estimates of the underlying distribution of durations.

For example, few, if any, reliability studies and prediction models are more comprehensive than those for developed for electronics components and assemblies for critical military applications. Even so, failure rates for similar items may vary by orders of magnitude even at the component level [Smith (2016)] introducing large differences between their reliability functions. The former military handbook MIL-HDBK-217F [MIL-HDBK-217F (1990)], long a standard for reliability prediction, warned:

> "Even when used in similar environments, the differences between system applications can be significant. . . . However, failure rates are also impacted by operational scenarios, operator characteristics,

maintenance practices, measurement techniques and differences of definition of failure. Hence, a reliability prediction should never be assumed to represent the expected field reliability as measured by the user (i.e., Mean-Time-Between-Maintenance, Mean-Time-Between-Removals, etc.). This does not negate its value as a reliability engineering tool; note that none of the applications discussed above requires the predicted reliability to match the field measurement."

An exception to the presumption that the observer has no knowledge of the reliability function is discussed in Section 4. That section deals with situations where event duration is measured in terms of active use, rather than dormant storage or standby conditions, and it applies as well to any other method for estimating remaining duration.

Several pro, con, and mixed commentaries on Gott's method have appeared in the literature based on well-reasoned statistical arguments; for example, see [Nature (1994)], [Monton (2006)], [Caves (2008)], [Pisaturo (2009)]. However, there seem to have been no documented attempts to implement and test the method systematically under conditions that would approximate the observation of events at no special time in their durations.

This paper begins to fill this gap by presenting the results of Monte Carlo simulations obtained using a variety of underlying distributions and mixtures of distributions. These tests are performed under conditions that both satisfy and violate the Copernican Principle of observation at no special time in the duration of the event.

This author would be grateful to be informed if and where the results of experiments testing Gott's method have been previously published so that proper attribution can be given in the future.

It would be a worthwhile future exercise to compare the results of reliability predictions for electronic or mechanical systems with wide variations on its input parameters to projections based on Gott's bounds, as well as perform retrospective comparisons to durations reported through real-world maintenance and repair experience.

3 Confidence Intervals Based on Conditional Survival with a Known Reliability Function

The value of Gott's method is that it can be applied when the observer has no knowledge of the history or underlying distribution that governs event duration. If such information is available, on the other hand, then it should be used.

If the reliability function $R(t)$ is known, then applying the method of conditional survival using the reliability function [Siewiorek (1982)] [Pham (2007)] [Hieke (2015)] [Hassani (2016)] can provide much better CIs than Gott's method.

The shorter term *conditional survival method* (or simply *conditional survival*) is used here in lieu of the full phrase "method of conditional survival using the reliability function."

Let $\Pr\{A \mid B\}$ denote the conditional probability of A, given that B occurs. Then the confidence C that an event is still ongoing at a time t, that is *subsequent* to observing that it is ongoing at t_{obs}, is found as follows:

$$\begin{aligned}
C &= \Pr\{\text{event is ongoing at } t \mid \text{event is ongoing at } t_{obs}\} \\
&= \frac{\Pr\{\text{event is ongoing at } t \cap \text{event is ongoing at } t_{obs}\}}{\Pr\{\text{event is ongoing at } t_{obs}\}} \\
&= \frac{\Pr\{\text{event is ongoing at } t\}}{\Pr\{\text{event is ongoing at } t_{obs}\}} \\
C &= \frac{R(t)}{R(t_{obs})} \\
&\text{or} \\
R(t) &= C \times R(t_{obs}) \ . \qquad (10)
\end{aligned}$$

The bounds of the two-sided conditional survival CI are found by setting C to two values as follows, and solving (10) for the time t in $R(t)$ that achieves those confidences. To obtain a 95% CI, (10) is solved for t where the LB is found by setting $C = 1 - \frac{1-0.950}{2} = 0.975$ and the UB is found by setting $C = \frac{1-0.950}{2} = 0.025$.

For some distributions, (10) may be solvable for t in closed form. Nevertheless, numerical solutions can be found easily for any specified $R(t)$ whether it is given as a mathematical expression or tabulated values.

The observer described in Section 2 was presumed to be dealing with a phenomenon for which the statistical history or underlying distribution is unknown, or even a one-time or one-of-a-kind event. A great deal more can be done by an informed observer with knowledge of the underlying distribution or behavior of a process. Such an observer may be able to use a technique such as the conditional survival method by being given the reliability function, approximating it based on prior observation of many such events, or using the physics or chemistry of the event to estimate remaining duration.

If the observer were to have external knowledge regarding the end of the event, such as a demolition order for a structure, then of course none of these calculations would be required or even relevant.

Examples of conditional survival solutions are given in this paper for CIs where $R(t)$ is approximated numerically by simulating the underlying distribution of the event durations 1,000,000 times, collecting the resulting durations, and using a lookup table to return t for a given value of $R(t)$ or return $R(t)$ for a given value of t. This approach is used for examples where $R(t)$ is based on a single distribution as well as cases where $R(t)$ is the result of a mixture of distributions.

4 Event Duration Determined by Active Use

This section applies to both Gott's method and the conditional survival method, as well as any other approach, and has to do with how event duration is measured.

When $R(t)$ or other detailed lifetime data are known, it must be specified as to whether t is measured in terms of operational hours (or some other accumulation of the stresses of active use) or strictly as calendar time [Smith (2016)].

Throughout this paper, the presumption is made that Gott's method is applied only when the underlying distribution for event duration is unknown to the observer. Nevertheless, it may be reasonable to admit a small amount of knowledge on the part of the observer in a situation where it is known or believed that equipment failures or other event terminations may occur far more frequently as a result of active operation than because of time spent in dormant storage or standby conditions.

Failure of the subject of an event can occur as a result of poor storage conditions or effects such as excessive cold or heat, freezing or thawing, humidity or aridity, evaporation or condensation, rusting, vibration, radiation, contamination, or infection. Items in storage can be lost to incidents such as earthquakes, floods, or vandalism. Failures also can occur even under the best of storage conditions due to mechanisms such as routine aging of materials.

Yet, even when lacking detailed knowledge of field failure rates,[12] or even simply their causes, it may be reasonable in many cases to assume that active use almost exclusively determines the operational lifetime of equipment. For this reason, reliability analyses often measure time in terms of that spent primarily or solely in active operation [MIL-HDBK-217F (1990)] [Smith (2016)].

For example, quantities of time measured as “operating hours,” “running hours,” “flying hours,” “engine hours,” “equipment hours,” “kilometers or

[12] As noted in Section 2, this could be the case even for well-studied technologies such as electronic components and assemblies.

miles driven," "throttle cycles," "number of operations or demands," "duty cycle," or "on/off cycles" may be far more crucial to count toward event duration than strict calendar time that includes storage, inactive, or standby hours [Smith (2016)]. In such cases it would be appropriate, if not essential, for any method that calculates an event's remaining duration t_{rem} or total duration t_{end}, on the basis of its observed past duration t_{obs}, to do so in terms of the appropriate measure of time when that quantity is known.

Where active operation is the major factor, this adjustment would make a significant difference for any method when such operation does not follow a regular or predictable pattern.

For example, consider an event duration defined as the failure-free performance of an item that has a lifetime determined by operational hours. Suppose that this item were used only intermittently and at very irregular intervals. In this example, let it be used continuously for a full week, then not again for a month, then again for a full day, and then sit dormant for five years. Assuming that storage or inactive time is not a significant factor in this example, then capturing the event's past duration t_{obs} as approximately 192 operational hours would be more informative than reporting the past duration as being about 45,000 hours of calendar time. Without a good estimate of the rate of future use, Gott's LB of $\frac{1}{39} \times 192 \approx 5$ operational hours would have better predictive value than calculating the LB as $\frac{1}{39} \times 45{,}000 \approx 1{,}150$ calendar hours.

On the other hand, this adjustment may not be necessary for Gott's method when active operation occurs on a regular or random basis even when active use is the primary determining factor for lifetime. This is demonstrated in the examples that follow in this section.

Suppose that, in contrast to the previous example of wildly intermittent and irregular use, the same item were in regular active operation for two hours per day, every work day, or just the first day of each month. In such cases, active time would clearly accumulate at a constant rate proportional with calendar time.

Even when active use does not occur on a regular basis, it may be sufficient for operational time to accumulate randomly at a relatively constant rate, particularly if the random occurrences of active use are statistically independent. This is demonstrated as follows.

Consider an example where active use of an item, for a relatively fixed number of hours, occurs on any given calendar day with the probability $p = 0.2$. If this occurrence is statistically independent from day to day, then the binomial distribution governs the number of active uses k that occur in n days.[13] In this example:

- In 100 days the expected number of active uses is 20, an exact 95% CI for the number of active uses is [11, 29], the length of this CI is 18, and the ratio of the length to the mean is 0.90.
- In 1,000 days the expected number of active uses is 200, the exact CI is [175, 226] with a length of 51, and the ratio of the length to the mean falls to about 0.26.
- In 10,000 days the expected number of active uses is 2,000, the exact CI is [1,921, 2,080] with a length of 159, and the ratio of the length to the mean is only about 0.08.

The same scaling process would work if this example were calculated on the basis of, say, five workdays per week instead of calendar days.

It can be seen in this example of randomly accumulating active operation that the expected number of active uses (as well as the total time spent in active use) would scale roughly proportionately with the calendar time. Furthermore, the difference between the LB and UB of the CI grows smaller, relative to the mean value, as the calendar time increases.

In these examples of both regular and random active use, calendar time would scale commensurately with an appropriate measure of duration based on active operation. Thus, Gott's method could measure the passage of time either as active use only or as strict calendar time, and those LBs and UBs on an event's remaining duration t_{rem} or total duration t_{end} would also scale proportionately.

[13] For the binomial distribution, the mean number of observed successes k in n trials, with success probability p on any given trial, is np. The LB and UB for observed values of k for the binomial distribution's exact CI can be obtained using the Excel® spreadsheet functions **BIN_LB_K_FOR_TRUE_P** and **BIN_UB_K_FOR_TRUE_P** given in Appendix C of [Debany (2023)].

These functions are written in the Microsoft® Excel® Visual Basic® for Applications or VBA™ language.

5 Simulations

Gott's method and the conditional survival method were tested using Monte Carlo simulations written in the GNU® C programming language.[14,15] Any of the results reported in this paper from the C program can be easily obtained independently from a spreadsheet implementation, although not to the same level of accuracy due to the smaller feasible number of events that can be simulated using rows in a spreadsheet.

High-quality random values were required for the simulations implemented in the C program. Uniformly distributed random numbers were produced using the **ran2** random number generator given by Press et al. [Press (2002)]. Those authors made this confident statement:

> "We think that, within the limits of its floating-point precision, **ran2** provides perfect random numbers; a practical definition of "perfect" is that we will pay $1,000 to the first reader who convinces us otherwise (by finding a statistical test that **ran2** fails in a nontrivial way, excluding the ordinary limitations of a machine's floating-point representation)."

It appears that there never was a successful claim against this offer. In any case, the offer of the **ran2** bounty was revoked by those authors in the 3^{rd} Edition (2007) of this reference, although reportedly not for a demonstrated failure of the random number generator.

---

[14] GNU® is a registered trademark of Free Software Foundation Corporation.

[15] The results presented in this paper were generated using a C program due to the large number of simulations as well as the large number of samples of the underlying distributions used to approximate the reliability functions.

However, the processes for event simulations were initially created and tested in spreadsheets that used only builtin functions and spreadsheet formulas with no programming required. They were used to double-check the logic and the outputs of the C program as it was developed.

Spreadsheets modeling 1,000 or even 10,000 event durations (and the independently generated $R(t)$ lookup table) updated in small numbers of seconds and were easily explored in summary and graphically.

The spreadsheet method was not used to produce the final reported results as it was not feasible to obtain the desired 1,000,000 observed events and samples of the reliability function. As shown by example in Section 6, the scaling problem for a spreadsheet is exacerbated by the fact that the Copernican Sample model must generate many more events than are eventually observed as ongoing.

As a precaution against potential initialization issues such as that famously suffered when the RC4 stream cipher was misused in the Wired Equivalent Privacy (WEP) of IEEE 802.11 [Edney (2004)], a "warm up" phase is used where the first 1,000 values generated by the random number generator are discarded after it is initialized with a random seed.

The term *random seed* means "the seed for the random number generator." The random seeds were generated systematically for each simulation (except where noted) to avoid duplication.

The exponential distribution as a function of t with parameter λ has the CDF $F(t) = 1 - e^{-\lambda\, t}$. Exponentially distributed random variates were obtained by using uniformly distributed values and the inverse of the CDF. With a uniformly distributed value u, the value $t = \frac{-\ln(1-u)}{\lambda}$ is exponentially distributed with a Mean Time To Failure (MTTF)[16] of $1/\lambda$ [Larson (1981)] [Viadinugroho (2021)].

Normally distributed random variates were obtained using the Box-Muller Transformation (BMT) [Weisstein (2019)] in the C program and using builtin functions in the spreadsheet implementations. The BMT takes two uniformly distributed values u and v to generate the value $z = \sqrt{-2\,\ln(u)}\,\sin(2\,\pi\,v)$, which is normally distributed as N[0, 1]. The value z is then scaled as $\sigma\,z + \mu$ to obtain a value normally distributed as N[μ, σ]. The sine function "sin()" used in the BMT takes its argument in radians and "N[μ, σ]" denotes the normal distribution with mean μ and standard deviation (SD) σ.

The Monte Carlo simulations implemented several distributions and mixtures of distributions. The primary model used for the evaluation of Gott's method was the Copernican Sample model as described in Section 2, but simulations based on the Copernican Observation model were run as well.

Table 1 lists the set numbers and definitions for the Monte Carlo simulations performed. Each set consists of 1,000,000 individual Monte Carlo simulations. The Copernican Sample model was used for Sets 1 through 25 and the Copernican Observation model was used for only Sets 1 through 9.

[16] The terms "time to failure" and MTTF are used in this paper in connection with the exponential distribution because of its historical applications in reliability analyses.

The conditional survival method requires knowing the underlying distribution or the reliability function $R(t)$ for event durations in order to solve (10) for the values of t that correspond to the LB and UB of the 95% CI using $C = 1 - \frac{1-0.950}{2} = 0.975$ and $C = \frac{1-0.950}{2} = 0.025$, respectively.[17] As noted in Section 3, $R(t)$ for the distributions or mixtures of distributions used in each Monte Carlo simulation was approximated numerically by simulating the underlying distribution of the event durations 1,000,000 times, storing those resulting durations in a lookup table, and using that table to find t that yields a given value of $R(t)$ or return $R(t)$ for a given value of t. The simulations used to approximate $R(t)$ were performed independently from those used for the Monte Carlo simulations. For consistency, this approach was used even when (10) could be solved for t in closed form.

The data presented in this paper are obtained from a subset of 86 sets of Monte Carlo simulations and the additional event duration simulations used to approximate $R(t)$. The 86 sets required a total run time of about nine minutes of wall clock time when run sequentially on a Intel® Core™ i3-7100U 2.40GHz processor.[18]

Even though 1,000,000 is a large number of Monte Carlo simulations by any reasonable standards, it would still be proper to give the actual or measured[19] confidences of the 95% CIs as binomial CIs themselves with LBs and UBs. Indeed, when the Monte Carlo simulations were run repeatedly with different random seeds, the estimates of the actual confidences did vary slightly.

However, when the lengths of the binomial CIs were calculated for the actual confidence levels as well as for their individual tail probabilities, the maximum length of any CI for the sets listed in Table 1 for the Copernican Sample and Copernican Observation models was 0.0012 in one case, with the next largest length being 0.0009. This means that the confidence levels and tail probabilities presented here would not vary except in their fourth decimal place. Thus, the confidences and tail probabilities given here are simply rounded to three decimal places.

[17] This is the last use of the word "respectively," as it should be clear in context that parallel lists of items are always given in the same order.

[18] Intel® Core™ are registered trademarks of Intel Corporation.

[19] "Actual" and "measured" are synonymous in this paper.

Table 2 through Table 7 are referenced in Section 6 and Section 7.

For the Monte Carlo simulations based on the Copernican Sample model, Table 2 gives the observed and true means of the event durations for each set of Monte Carlo simulations. The *true mean* is the average duration of all generated events regardless of whether or not they are observed, and is obtained from the simulation data; the true mean will vary slightly from the ideal (or theoretical) mean as specified for that distribution. The *observed mean* is the average duration of only the events captured at the sample time.

Table 3 gives the actual 2.5%ile LB and 97.5%ile UB for ρ for Gott's method calculated as (1) in Section 2.[20]

Table 4 gives the actual confidences and tail probabilities for Gott's projected t_{end} as well as those obtained using the conditional survival method.

Similarly, Table 5, Table 6, and Table 7 give these statistics for the Copernican Observation model.

These statistics based on Monte Carlo simulations of event durations can be considered to be derived from a total of 1,000,000 captured events from cross sections of one or more populations, each consisting of one or more events.

[20] The notation "%ile" denotes "percentile."

Table 1: Key for sets of Monte Carlo simulations. All sets are performed for the **Copernican Sample** model, but only Sets 1 through 9 are performed for the **Copernican Observation** model. Set numbers are the same for both models.

Set	**Definition**
1	Uniform on [0, 20]
2	Exponential with MTTF = $1/\lambda = 10$
3	Normal with $\mu = 10$, $\sigma = 5$
4	Normal with $\mu = 10$, $\sigma = 2$
5	Normal with $\mu = 20$, $\sigma = 0.01$
6	Exponential with MTTF uniformly distributed between $\text{MTTF}_{lo} = 5$ and $\text{MTTF}_{hi} = 15$
7	Exponential with λ uniformly distributed between $\lambda_{lo} = 0.0667$ and $\lambda_{hi} = 0.2000$
8	Uniform on [0, z] where z is normally distributed with $\mu = 20$, $\sigma = 5$
9	SSA lifetimes
10	Uniform on [0, 20], Baby Boom start times normal with $\mu_{BB} = 60$, $\sigma_{BB} = 20$
11	Uniform on [0, 20], Baby Boom start times normal with $\mu_{BB} = 80$, $\sigma_{BB} = 20$
12	Uniform on [0, 20], Baby Boom start times normal with $\mu_{BB} = 100$, $\sigma_{BB} = 20$
13	Uniform on [0, 20], Baby Boom start times normal with $\mu_{BB} = 120$, $\sigma_{BB} = 20$
14	Uniform on [0, 20], Baby Boom start times normal with $\mu_{BB} = 140$, $\sigma_{BB} = 20$
15	Exponential with MTTF = $1/\lambda = 10$, Baby Boom start times normal with $\mu_{BB} = 60$, $\sigma_{BB} = 20$
16	Exponential with MTTF = $1/\lambda = 10$, Baby Boom start times normal with $\mu_{BB} = 80$, $\sigma_{BB} = 20$
17	Exponential with MTTF = $1/\lambda = 10$, Baby Boom start times normal with $\mu_{BB} = 100$, $\sigma_{BB} = 20$
18	Exponential with MTTF = $1/\lambda = 10$, Baby Boom start times normal with $\mu_{BB} = 120$, $\sigma_{BB} = 20$
19	Exponential with MTTF = $1/\lambda = 10$, Baby Boom start times normal with $\mu_{BB} = 140$, $\sigma_{BB} = 20$
20	Normal with $\mu = 10$, $\sigma = 2$, Baby Boom start times normal with $\mu_{BB} = 60$, $\sigma_{BB} = 20$
21	Normal with $\mu = 10$, $\sigma = 2$, Baby Boom start times normal with $\mu_{BB} = 80$, $\sigma_{BB} = 20$
22	Normal with $\mu = 10$, $\sigma = 2$, Baby Boom start times normal with $\mu_{BB} = 100$, $\sigma_{BB} = 20$
23	Normal with $\mu = 10$, $\sigma = 2$, Baby Boom start times normal with $\mu_{BB} = 120$, $\sigma_{BB} = 20$
24	Normal with $\mu = 10$, $\sigma = 2$, Baby Boom start times normal with $\mu_{BB} = 140$, $\sigma_{BB} = 20$
25	SSA lifetimes, Baby Boom start times generated using CDC et al. birth data

Table 2: Observed and True Mean for event durations or lifetimes for the **Copernican Sample** model.

Set	**Observed Mean**	**True Mean**
1	13.329	10.003
2	19.996	9.999
3	12.433	10.275
4	10.397	9.999
5	20.000	20.000
6	21.621	9.991
7	18.186	8.238
8	14.163	10.001
9	78.737	74.629
10	14.267	10.001
11	13.674	10.000
12	13.131	10.000
13	12.651	10.001
14	12.264	10.000
15	25.981	10.002
16	20.504	9.999
17	17.453	10.000
18	15.666	9.999
19	14.511	10.000
20	10.580	10.000
21	10.463	10.000
22	10.371	10.000
23	10.286	10.000
24	10.219	10.000
25	78.354	74.624

Table 3: Actual Gott LB and UB ratios ρ as defined by (1) for 95% CIs for the **Copernican Sample** model.

Set	**Actual LB**	**Actual UB**
1	0.026	38.823
2	0.026	39.203
3	0.026	39.029
4	0.026	38.727
5	0.026	38.823
6	0.026	38.926
7	0.026	38.764
8	0.026	39.076
9	0.026	39.401
10	0.016	20.380
11	0.022	30.372
12	0.030	42.826
13	0.041	56.260
14	0.056	72.357
15	0.016	16.105
16	0.024	29.919
17	0.034	46.187
18	0.045	63.838
19	0.057	82.477
20	0.017	23.384
21	0.022	31.295
22	0.028	41.178
23	0.037	52.255
24	0.049	63.369
25	0.025	50.000

Table 4: Gott and conditional survival confidence levels (the probability of the remaining duration being between the LB and UB) and tail probabilities (the probabilities of the remaining duration being below the LB or above the UB) for the **Copernican Sample** model.

Set	**Prob Within Gott's LB & UB**	**Prob Below Gott's LB**	**Prob Above Gott's UB**	**Prob Within Cond. Surv. LB & UB**	**Prob Below Cond. Surv. LB**	**Prob Above Cond. Surv. UB**
1	0.950	0.025	0.025	0.950	0.025	0.025
2	0.950	0.025	0.025	0.950	0.025	0.025
3	0.950	0.025	0.025	0.950	0.025	0.025
4	0.950	0.025	0.025	0.950	0.025	0.025
5	0.950	0.025	0.025	0.950	0.025	0.025
6	0.950	0.025	0.025	0.950	0.025	0.026
7	0.950	0.025	0.025	0.950	0.025	0.025
8	0.950	0.025	0.025	0.950	0.025	0.025
9	0.950	0.025	0.025	0.960	0.019	0.021
10	0.947	0.040	0.013	0.950	0.025	0.025
11	0.951	0.029	0.020	0.950	0.025	0.025
12	0.952	0.021	0.027	0.950	0.025	0.025
13	0.948	0.016	0.036	0.950	0.025	0.025
14	0.943	0.012	0.045	0.950	0.025	0.025
15	0.950	0.040	0.010	0.950	0.025	0.025
16	0.954	0.026	0.019	0.950	0.025	0.025
17	0.952	0.019	0.030	0.950	0.025	0.025
18	0.945	0.014	0.041	0.950	0.025	0.025
19	0.937	0.011	0.052	0.950	0.025	0.025
20	0.948	0.037	0.015	0.950	0.025	0.025
21	0.951	0.029	0.020	0.950	0.025	0.025
22	0.951	0.023	0.026	0.950	0.025	0.025
23	0.949	0.018	0.033	0.950	0.025	0.025
24	0.946	0.013	0.040	0.950	0.025	0.025
25	0.942	0.026	0.032	0.949	0.031	0.021

Table 5: Observed and True Mean for event durations or lifetimes for the **Copernican Observation** model. The two means are the same, unlike the case for the Copernican Sample model.

Set	Observed and True Mean
1	9.997
2	9.999
3	10.276
4	10.001
5	20.000
6	10.010
7	8.242
8	10.000
9	74.604

Table 6: Actual Gott LB and UB ratios ρ as defined by (1) for 95% CIs for the **Copernican Observation** model.

Set	Actual LB	Actual UB
1	0.026	39.126
2	0.026	38.748
3	0.026	39.065
4	0.026	38.815
5	0.026	38.990
6	0.026	38.991
7	0.026	39.145
8	0.026	39.135
9	0.026	39.236

Table 7: Gott and conditional survival confidence levels and tail probabilities for the **Copernican Observation** model.

Set	Prob Within Gott's LB & UB	Prob Below Gott's LB	Prob Above Gott's UB	Prob Within Cond. Surv. LB & UB	Prob Below Cond. Surv. LB	Prob Above Cond. Surv. UB
1	0.950	0.025	0.025	0.950	0.025	0.025
2	0.950	0.025	0.025	0.950	0.025	0.025
3	0.950	0.025	0.025	0.950	0.025	0.025
4	0.950	0.025	0.025	0.950	0.025	0.025
5	0.950	0.025	0.025	0.951	0.025	0.025
6	0.950	0.025	0.025	0.950	0.025	0.025
7	0.950	0.025	0.025	0.950	0.025	0.025
8	0.950	0.025	0.025	0.950	0.025	0.025
9	0.950	0.025	0.025	0.956	0.024	0.020

6 Constant Populations of Events

As outlined in Section 2, the Copernican Principle is the assumption that, if an ongoing event is observed, the capture time of that event is uniformly distributed in its duration. For the purpose of constructing the Monte Carlo simulations used here, the Copernican Sample model defined in Section 2 implements this principle using *constant populations* of events with *uniformly distributed start times* within a finite timespan, where the sample time occurs at a *fixed point* in that timespan.

Each simulation represents a single event within either a population of multiple events or a population consisting of just that one event.

The event durations are generated from a given distribution or mixture of distributions. For example, Set 1, as noted in its definition in Table 1, tests the Copernican Sample model for events with durations uniformly distributed between 0 and 20, or U[0, 20]. Figure 1 shows 100 such events with start times uniformly distributed in a simulation calendar timespan from 0 to 120 and a fixed simulation calendar sample time at 100; the unit of time is arbitrary and is ignored. The sample time is selected to be much larger than any likely event duration so that the population can settle into what amounts to a constant or steady state.[21,22,23]

Figure 1 is an example of sampling a cross section of a population. If the generated event does not intersect with the sample time, then that event is not observed. The observer therefore captures either no event or a *single* event when sampling a population; events are generated until the required number are captured. When an event is captured, the observer has only one value of the observation time t_{obs} on which to base the CI for either t_{rem} or t_{end}.

[21] See footnote #5 regarding calendar time and simulation time.

[22] Any start times that occur after the sample time result in events that can never be sampled. For this reason, for the purposes of these simulations the choice of the (finite) timespan is arbitrary as long as it is larger than the sample time. Increasing the timespan has no effect other than to require more events to be generated in order to observe the specified quota of 1,000,000 events at the sample time.

[23] Some distributions, such as the exponential or normal distributions, can generate arbitrarily large durations but with very low probability. The sample time was chosen to be considerably larger than any reasonable duration that would be generated.

Figure 1 shows 100 events belonging to one or more populations, but it is seen that only 11 of the events are ongoing when sampled at the simulation calendar time of 100. For each set of Monte Carlo experiments reported, simulations representing individual events are created until 1,000,000 events have been observed as ongoing, and the statistics reported in Table 2, Table 3, and Table 4 are calculated.

For example, for Set 1 it was necessary to create 6,000,000 events to observe the quota of 1,000,000 ongoing events. Even though many more events had to be generated than the eventual number of event durations that intersected with the sample time, this constituted "1,000,000 Monte Carlo simulations" because only the number of sampled events mattered. In spite of the large number of events generated, the C program that implemented these simulations for the Copernican Sample model for Set 1 ran for only 2.8 seconds of wall clock time, which included the time for the independent set of simulations used to approximate $R(t)$ for the underlying distribution U[0, 20].

The constant or steady state populations simulated for this work correspond to Sets 1 through 9 defined in Table 1.

Set 1, using event durations uniformly distributed as U[0, 20], has already been described in this section.

Figure 2 shows 100 events typical of Set 2, which uses events having exponentially distributed durations with an MTTF of $1/\lambda = 10$,

Figure 3 shows 100 events typical of Set 3, which uses events having normally distributed durations with mean $\mu = 10$ and standard deviation $\sigma = 5$, or N[10, 5].

Because an event duration must be greater than 0, any generated duration with a negative value is discarded and another random variate is generated.[24] For this reason, the "normally distributed" durations used in this work are not truly normal but are truncated to some degree. Set 3 had a relatively large standard deviation compared to its mean, resulting in about 2.3% of the generated durations being negative and thus discarded.

[24] Any event generated with a duration of exactly 0, if that ever happened, would be discarded automatically because it could never be sampled.

Set 4 is for events having durations normally distributed as N[10, 2]. Set 5 is for events having durations normally distributed as N[20, 0.01]; with such a small standard deviation it was practically equivalent to a constant duration of 20. Neither set would be likely to generate even a single negative duration in the experiments described here.

Set 6 uses a mixture of exponential distributions. For each Monte Carlo simulation in this set, an MTTF is chosen from the uniform distribution U[5, 15], and then an event with an exponentially distributed duration with that MTTF is generated.

Set 7 uses the same approach as for Set 6 except that, for each Monte Carlo simulation in this set, a value of λ is chosen from the uniform distribution U[$\frac{1}{15}$, $\frac{1}{5}$], and then an event with an exponentially distributed duration with that MTTF $= 1/\lambda$ is generated.

Set 8 uses a mixture of uniform distributions. For each such random variate, a truncated normally distributed value $z \geq 0$ is generated from N[20, 5] (about 0.003% of the values were negative and thus discarded), and then the duration of an event is chosen from the uniform distribution U[0, z]; this distribution is written here as U[0, N[20, 5]]. Figure 4 shows 100 such events.

Set 9 creates events with durations that model human lifetimes. Figure 5 shows $R(t)$ for male lifetimes in the United States based on Social Security Administration (SSA) data [SSA (2023)]. Figure 6 shows 100 such events with start times uniformly distributed in a simulation calendar timespan from 0 to 600 and a fixed simulation calendar sample time at 500.

The well-known *Inspection Paradox* arises when an observer samples ongoing phenomena; for example, see [Heyman (1982)] [Downey (2019)]. A commonly used scenario is where the goal is to determine the lifetime of a type of light bulb by surveying those in operation in some facility. An observer arrives at the site of a light bulb that is currently operational, inquires how long it has been so, and then records its total duration when it burns out.

In this scenario, based on the data for many such samples, the observer estimates the MTTF for the light bulbs by dividing the sum of the recorded total durations by the number of light bulbs surveyed.[25]

If the lifetimes are exponentially distributed, then this type of survey will result in an estimate that is double the true MTTF: hence the "Paradox." The reason for this is that such an experiment has a greater chance of observing an event of longer duration than an event of shorter duration.

For a distribution with time to failure with mean μ and standard deviation σ [Heyman (1982)], the estimated MTTF that would result from such a sampling protocol is

$$\widehat{\mathrm{MTTF}} = \mu + \frac{\sigma^2}{\mu} \quad . \tag{11}$$

As noted in Section 5, Table 2 gives the observed and true mean event durations for each set of Monte Carlo simulations for the Copernican Sample model. Recall that the observer samples only a *single* event from the population to calculate a CI, but these are the mean values for event durations that would have been obtained from *multiple samples*.

Set 1 had durations uniformly distributed as U[0, 20]. A uniform distribution U[a, b] has an expected value $\mu = \frac{a+b}{2}$ and standard deviation $\sigma = \sqrt{\frac{(b-a)^2}{12}}$ [Rohatgi (1979)]. The true mean of the durations generated for Set 1, including those durations not observed because they did not intersect with the sample time, was 10.0025 which closely matches the ideal MTTF. The observed mean, however, was 13.3286; when the values of μ and σ for the uniform distribution are substituted into (11), the result is $\widehat{\mathrm{MTTF}} = 13.3333$ which is extremely close to this observed mean duration.

[25] Conducting a light bulb survey in a facility subject to shutdown is an example of where it could be appropriate to count operational time instead of strict calendar time as outlined in Section 4. Suppose that, during the course of the survey, the facility were initially in a phase of normal activity where the light bulbs experienced steady duty cycles, followed by a long shutdown period, and then reopened for normal activity.

If the environmental conditions in the facility during the shutdown were favorable and not destructive to those items, then it would be reasonable to assume that the dormant period during which the light bulbs were turned off should not be tallied in the event durations captured in the survey.

Regardless of the issue of the Inspection Paradox, this example again points out the distinction between operational lifetime and calendar lifetime.

Set 2 had durations exponentially distributed with MTTF $= 1/\lambda = 10$. An exponential distribution with expected value $1/\lambda$ has standard deviation $1/\lambda$ as well. The true mean of the durations generated for Set 2, including those not observed, was 9.9992 which again closely matches the ideal MTTF. The observed mean was 19.9956, however, and when the values of the mean and standard deviation for this exponential distribution are substituted into (11), the predicted observed mean is $\widehat{\text{MTTF}} = 20.0000$, which is extremely close to this observed mean duration.

Sets 3, 4, and 5 had durations based on truncated normal distributions with various combinations of μ and σ.

As noted earlier, Set 3 had a significant number of generated durations that were negative and were discarded. As a result, its true mean duration was 10.2748 which was larger than the ideal mean of 10. Even so, the predicted observed mean obtained from (11) was $\widehat{\text{MTTF}} = 12.5000$, which is quite close to the observed mean of 12.4328.

Set 4 had a true mean duration of 9.9994 which was very close to the ideal mean of 10. The predicted observed mean obtained from (11) was $\widehat{\text{MTTF}} = 10.4000$, which is extremely close to the observed mean of 10.3970.

Set 5 was practically equivalent to a constant duration of 20 and its true mean duration was indeed found to be 20.0000. The predicted observed mean obtained from (11) was $\widehat{\text{MTTF}} = 20.0000$, as was the observed mean.

These examples demonstrate that the Inspection Paradox operates as expected for the Copernican Sample model.

Having established this fact, however, the non-intuitive results of the Inspection Paradox in turn open the possibility that sampling events in this manner might not yield uniformly distributed observation times within event durations: if samples taken in this way tend to capture longer durations, then might they also tend to occur later in an event duration? It is fair to expend effort to provide evidence that this is not the case and the assumption of the Copernican Principle does indeed hold when events are generated and captured using the Copernican Sample model described in this paper.

As noted in Section 2, this is equivalent to asking if $\frac{t_{obs}}{t_{end}}$ is uniformly distributed on the interval [0, 1] under the Copernican Sample model.

This question is addressed by the pairs of plots in Figure 7 through Figure 16, which correspond to Sets 1, 2, 4, 8, and 9.

The first plot in each pair shows a crude histogram of the frequencies of the ratios $\frac{t_{obs}}{t_{end}}$ for 100 bins in the interval [0, 1] for the 1,000,000 Monte Carlo simulations performed for each set.

The second in each pair is a scatter plot of $\frac{t_{obs}}{t_{end}}$ as a function of t_{end} for the first 5,000 of the Monte Carlo simulations performed for each set. The smaller number of data points used was chosen for readability of the plots.

A perfectly even distribution of $\frac{t_{obs}}{t_{end}}$ would have 0.01 in each of the 100 frequency bins of a histogram. There are slight deviations above and below the value of 0.01 but, being based on 1,000,000 samples, the frequencies are so close to the ideal distribution that it should be unnecessary to do a test such as χ^2 to agree that the Copernican Principle might hold for each of these experiments.[26]

The word "might" is used above because the even distribution of $\frac{t_{obs}}{t_{end}}$ taken over all durations is necessary, but not sufficient, to show that this ratio is evenly distributed at *every* duration. In lieu of showing multiple histograms of this ratio taken at small bands of durations, a qualitative demonstration employs a second plot in each pair that is a scatter plot of the density of values of this ratio over all durations.

Figure 8 shows such a plot for durations uniformly distributed as U[0, 20]. This plot shows increasing density of points from left to right on the Duration axis, as would be expected due to events with longer durations being more likely to be observed. The density of points on the Ratio axis, however,

[26] Even so, a χ^2 test for the actual counts in the 100 bins used to plot the frequencies in the Set 1 histogram shown in Figure 7, compared to their expected uniform distribution counts of 10,000 out of 1,000,000 samples of $\frac{t_{obs}}{t_{end}}$, yields a p-value of 0.5919.

As this p-value is larger than a level of significance of 0.05 (see footnote #8), one cannot reject the null hypothesis that the deviation from the expected values is due to chance alone, and may informally "accept" that these ratios are uniformly distributed.

The χ^2 tests for the bins obtained for Sets 2, 4, 8, and 9 (yielding p-values 0.3768, 0.4608, 0.7739, and 0.1967) also indicate that these ratios may be uniformly distributed.

appears to the eye to be uniform along that axis and suggests that t_{obs} occurs uniformly within the interval [0, t_{end}] regardless of the event duration.

Each pair of plots in Figure 7 through Figure 16 shows the same property that the density of the points on the Ratio axis appears to the eye to be uniform along that axis. This is a strong indication that the Copernican Principle does indeed hold for the Copernican Sample model.

Now it remains to be seen if Gott's method yields the target 95% CIs. This demonstration is approached from complementary directions that are equivalent but both informative.

The first part of the demonstration extracts Gott's bounds from the CIs. Based on 1,000,000 Monte Carlo simulations in each set, Table 3 gives the actual 2.5%ile LB and 97.5%ile UB on ρ for Gott's method. Recalling that ρ is the ratio of the remaining duration following the observation time, as given in (1) in Section 2, these measured values of ρ should closely match Gott's calculated LB and UB as given by (4).

And indeed, for Sets 1 through 9, being for constant populations, the actual LB is 0.026 in each case, which would be the rounded value of the LB of 1/39 predicted by (4). The actual UB, if rounded to an integer, would be 39 as predicted by (4) as well.

This has shown that the measured LB and UB on ρ match Gott's calculated bounds for a variety of distributions and mixtures of distributions.

The second part of the demonstration evaluates these Monte Carlo simulation data from the opposite direction. Table 4 gives the probability or confidence that the measured ρ for a single Monte Carlo simulation falls within the calculated LB and UB; these confidences are based again on the 1,000,000 Monte Carlo simulations for each set. For the constant populations tested in Sets 1 through 9, the confidence levels shown in the second column are the target value of 0.950 when rounded to three digits.

This has demonstrated that Gott's method yields 95% CIs as intended when the Copernican Principle holds, regardless of the underlying distribution and the observer having no knowledge of that distribution.

The third and fourth columns of Table 4 give the tail probabilities for the CIs. In each case for Sets 1 through 9, the tail probabilities are 0.025 when rounded, indicating that the CIs are central as well as providing the target 95% confidence.

Table 4 also presents the results of testing the conditional survival method based on (10) that uses full knowledge of the underlying distribution, the reliability function $R(t)$.[27] The fifth column of Table 4 gives the probability or confidence that the measured ρ for a single Monte Carlo simulation falls within the calculated conditional survival LB and UB, based on the 1,000,000 Monte Carlo simulations for each set. For the constant populations tested in Sets 1 through 9, the confidence levels shown in the second column are the target value of 0.950 when rounded, except for a slight difference in the case of Set 9 which was 0.960.

The sixth and seventh columns of Table 4 give the tail probabilities for the conditional survival CIs, and the tail probabilities are 0.025 or 0.026 when rounded, varying significantly only in the case of Set 9, indicating that the conditional survival CIs also appear to be central.

As noted in Section 2, when the underlying reliability function is known to the observer, the conditional survival CIs have much smaller lengths than those obtained from Gott's method. This is demonstrated in the following examples taken from two of the sets of Monte Carlo simulations.

Consider some values taken from the simulations for Set 1, which uses the uniform distribution U[0, 20]. For $t_{obs} = 2.0863$, Gott's LB and UB on the remaining duration t_{rem} are 2.1398 and 83.4533 and the conditional survival LB and UB on t_{rem} are 2.5353 and 19.5458; the length of a CI is the difference between the LB and UB, so the length using Gott's method is 81.3135 and the length using the conditional survival method is 17.0105.

Continuing with Set 1, for $t_{obs} = 12.6895$, Gott's LB and UB on t_{rem} are 13.0149 and 507.5793 and the conditional survival LB and UB are 12.8683 and 19.8132; their lengths are 494.5644 and 6.9449.

[27] As described in Section 3 and Section 5, $R(t)$ is approximated numerically in this study by capturing 1,000,000 samples of the distribution independently of the Monte Carlo simulations.

Similarly, consider some values taken from the simulations for Set 3, which uses the truncated normal distribution N[10, 5]. For $t_{obs} = 5.1894$, Gott's LB and UB on t_{rem} are 5.3225 and 207.5773 and the conditional survival LB and UB are 5.5917 and 20.1817; their lengths are 202.2548 and 14.59.

Continuing with Set 3, for $t_{obs} = 10.3922$, Gott's LB and UB on t_{rem} are 10.6587 and 415.6881 and the conditional survival LB and UB are 10.5387 and 21.3391; their lengths are 405.0294 and 10.8004.

In these examples, the lengths for Gott's method are seen to be much larger than those for the conditional survival method due to the UBs for Gott's method being much larger. But it is also seen that the LBs obtained from Gott's method and the conditional survival method are not very different.

As noted in the Introduction, the LB can be a measure of trustworthiness when the "important action" is the condition of the event duration not ending. It would be a fortunate happenstance if, as suggested in these examples, that the LBs obtained from Gott's method that uses no knowledge of the underlying distribution are often close to those of the conditional survival method that uses effectively perfect knowledge of the underlying distribution.

This close conjunction of the LBs was explored further. The plots in Figure 17 through Figure 21 correspond to Sets 1, 2, 4, 8, and 9. These figures show plots of the conditional survival LB as a function of Gott's LB for 10,000 Monte Carlo simulations for these sets with constant populations. The red line in each plot would be equality with Gott's LBs.

The LBs for Gott's method and the conditional survival method do not differ greatly in these cases. Three of these plots show these LBs practically coinciding within the visible resolution, while the other two plots show differences for the smaller LBs but still closely coincide for the larger LBs. Where there are large differences in these cases, Gott's LBs are smaller than the conditional survival LBs, and so they conservatively project smaller remaining durations.

The major value of Gott's method is that the underlying distribution does not have to be known to the observer if the Copernican Principle can be assumed to hold for the time of observation. That Gott's LBs were uncannily close to the conditional survival LBs in these examples of two-sided 95% CIs may be an additional benefit of Gott's method for the estimation of trustworthiness. It would remain for statisticians to determine how closely, in general, Gott's LB might approximate the LBs obtainable from other methods that use detailed knowledge about the underlying distribution of event durations.

One-sided CIs were not explored to any great extent in this work, and future investigations could also determine how closely the 95% LBs resulting from (8) tend to match those obtained from other methods.

As discussed in Section 2, the Copernican Sample model appears to be the most reasonable in practice and is the one emphasized in this work. The alternative Copernican Observation model is now addressed even though it is immediately deprecated because applying it to an actual situation would appear to be unrealistic. To review, the Copernican Observation model is where the observation time is chosen *explicitly* so that t_{obs} is uniformly distributed on the interval $[0,\ t_{end}]$; this is in contrast to the Copernican Sample method that has been shown in this section to *implicitly* select an observation time uniformly distributed over the duration of the event as a consequence of the sampling within a timespan that may, or may not, capture an ongoing event.

It would seem that there should be no difference between the Copernican Sample model and the Copernican Observation model as they both involve a uniform distribution of the observation time within the event duration. The two models have different behaviors, however, although Gott's CIs are found to be valid for the Copernican Observation model as well.

Table 5 for the Copernican Observation model corresponds to Table 2 for the Copernican Sample model. Table 5 consists of a single column because the observed and true mean of the Monte Carlo simulation durations are the same: there is no Inspection Paradox in this model.[28] The true means

[28] For the example of the light bulb survey discussed earlier in this section, in the Copernican Observation model *all* light bulbs would be observed from the time of installation, not just the ones still in operation when the observer arrives.

obtained in the Copernican Observation simulations are practically the same as the true means obtained in the Copernican Sample simulations, which would be expected as a result of running 1,000,000 Monte Carlo simulations for each set.

Table 6 for the Copernican Observation model corresponds to Table 3 for the Copernican Sample model. The actual LB and UB on ρ for the Copernican Observation model again closely match the calculated values for Gott's method given as (4).

Table 7 for the Copernican Observation model corresponds to Table 4 for the Copernican Sample model. The second, third, and fourth columns show that the CI confidence and tail probabilities are correct for central CIs for Gott's method.

The fifth, sixth, and seventh columns of Table 7 give the confidence and tail probabilities that the measured ρ for a single Monte Carlo simulation falls within the calculated conditional survival LB and UB, based on the 1,000,000 Monte Carlo simulations for each set. These values again closely match the target confidence of 95% and tail probabilities of 2.5% so the CIs are central.

Aside from the fact that applying the Copernican Observation model in the real world would be problematic, another factor regarding the conditional survival approach suggests that the Copernican Observation model is less realistic than the Copernican Sample model.

For a given value of t_{obs}, the conditional survival expression (10) obtained in Section 3 requires the value of $R(t_{obs})$. For the Copernican Sample model, $R(t_{obs})$ for a sampled event is obtained by finding the value of $R(t)$ for which $t = t_{obs}$. For the Copernican Observation model, however, the CIs require $R(t_{obs})$ to be 1 for their target confidence level to match the actual confidence level from the simulations, *regardless* of the value of t_{obs}. Thus, (10) for the Copernican Observation model is independent of the observation time and reduces to simply $R(t) = C$.

But it is well-known and easily shown that the conditional probability of surviving for a period of time beyond a given observation time does indeed depend on that observation time for an event duration governed by any distribution other than the memoryless exponential distribution.[29] In any case, the conditional survival LB and UB for the 95% CI in the Copernican Observation model are found by solving the reliability function for t in the reduced expression $R(t) = C$ using $C = 0.975$ and $C = 0.025$, as outlined in Section 3.

Earlier in this section, Figure 8 showed a plot of $\frac{t_{obs}}{t_{end}}$ for durations uniformly distributed as U[0, 20] for the Copernican Sample model. That figure showed apparently uniform density across the Ratio axis, but the density of points increased from left to right on the Duration axis due to events with longer durations being more likely to be observed. In contrast to those results, Figure 22 shows the same situation for simulations using the Copernican Observation model, where the observation of an event occurs explicitly at a randomly selected point in time uniformly distributed over the event duration. To the eye, this figure shows apparently uniform density of points across the Duration axis as well as across the Ratio axis. The corresponding plots for the other distributions modeled here show similar differences in density between the Copernican Sample and Copernican Observation models, although it is most clearly seen in the case of events with durations that are uniformly distributed.

[29] For example, see [Hieke (2015)].

This property of the conditional probability of survival can be demonstrated with a simple example outside of the discussion of the Copernican Principle.

Consider an event with a duration known to be uniformly distributed as U[0, 100]. For this distribution, the probability that the event survives to time t is given by $R(t) = 1 - \frac{t}{100}$ for $0 \leq t \leq 100$. The probability of the event surviving to time $t = 20$ is $R(20) = 0.8000$. Independently, the probability of surviving to time $t = 21$ would be $R(21) = 0.7900$ which is slightly less than at $t = 20$.

However, it is consistent with everyday experience that an event that is ongoing at a certain time is highly likely still to be ongoing a very short time later. Indeed, using (10), the conditional probability of surviving for one more unit of time, given that it was observed as ongoing at $t = 20$, would be $\frac{R(21)}{R(20)} = 0.9875$ which is considerably higher than the independent probability of surviving to $t = 21$.

For the same distribution, this is shown even more dramatically at time $t = 80$, for which the probability of the event surviving is $R(80) = 0.2000$. Independently, the probability of surviving to time $t = 81$ would be $R(81) = 0.1900$ which is slightly less than at $t = 80$. Applying (10), the conditional probability of surviving for one more unit of time, given that it was observed as ongoing at $t = 80$, would be $\frac{R(81)}{R(80)} = 0.9500$; again, this is far higher than the independent probability of surviving to $t = 81$.

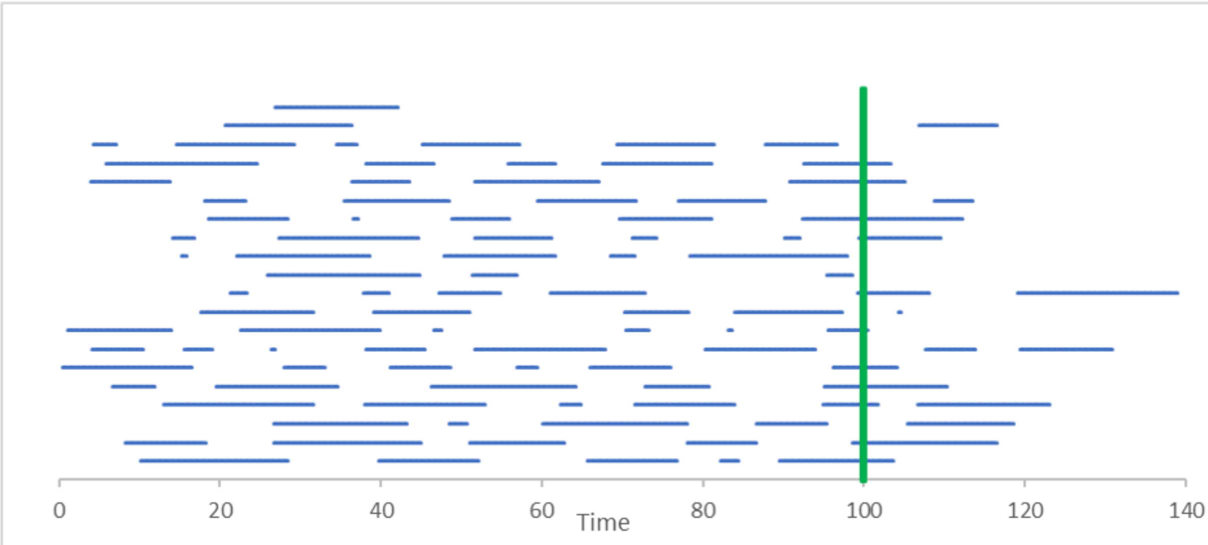

Figure 1: Example of 100 event durations typical of Set 1, distributed as U[0, 20], constant population start times uniformly distributed in a simulation calendar timespan from 0 to 120. Sample is taken at simulation calendar time 100.

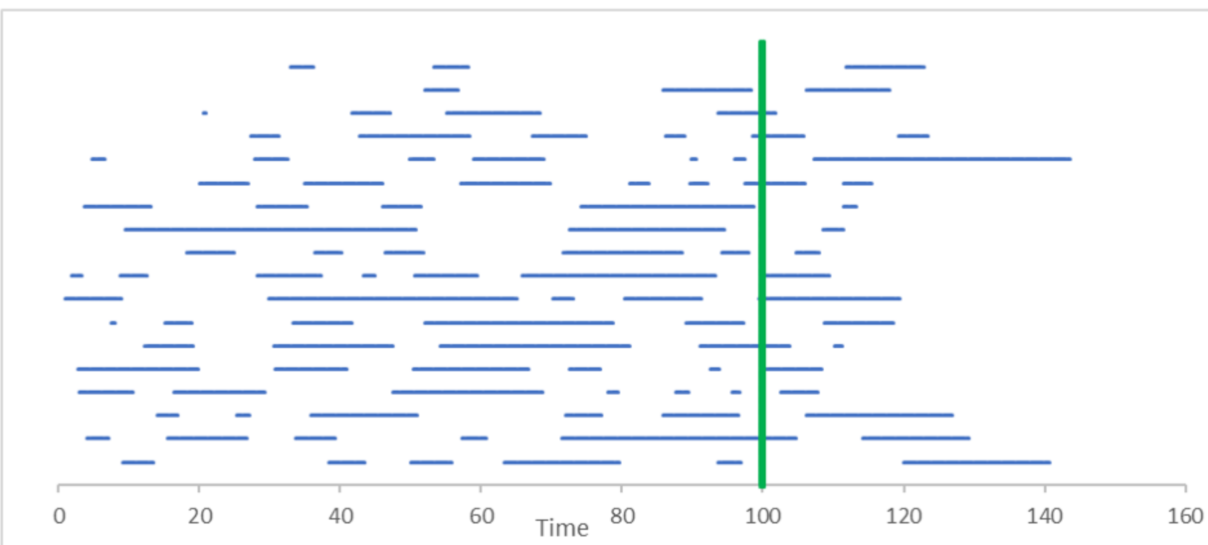

Figure 2: Example of 100 event durations typical of Set 2, exponentially distributed with MTTF = $1/\lambda = 10$, constant population start times uniformly distributed in a timespan from 0 to 120. Sample time is at 100.

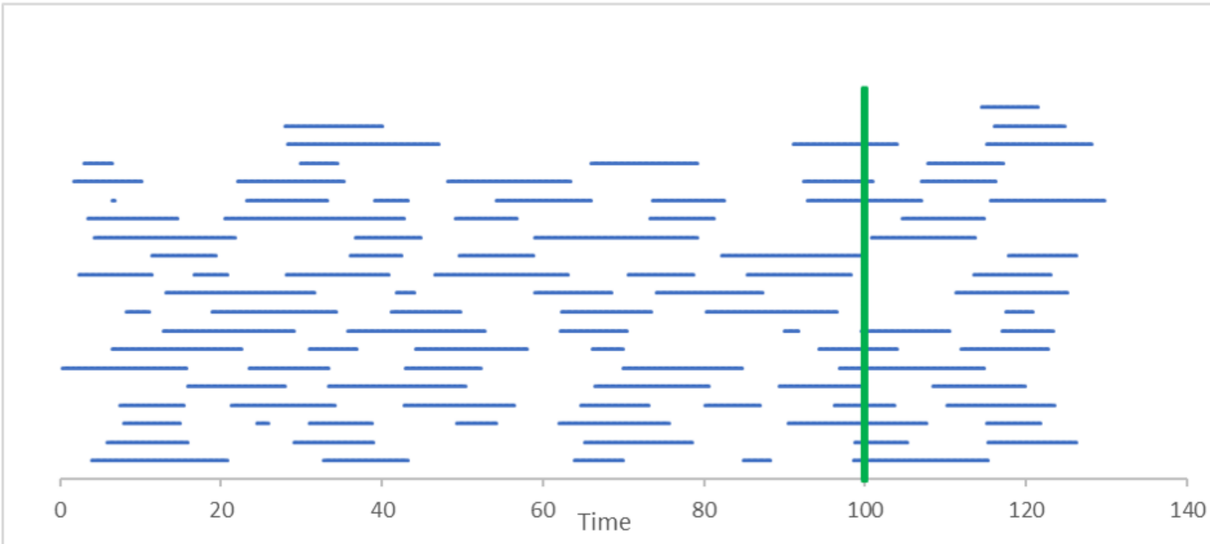

Figure 3: Example of 100 event durations typical of Set 3, distributed as N[10, 5], constant population start times uniformly distributed in a timespan from 0 to 120. Sample time is at 100.

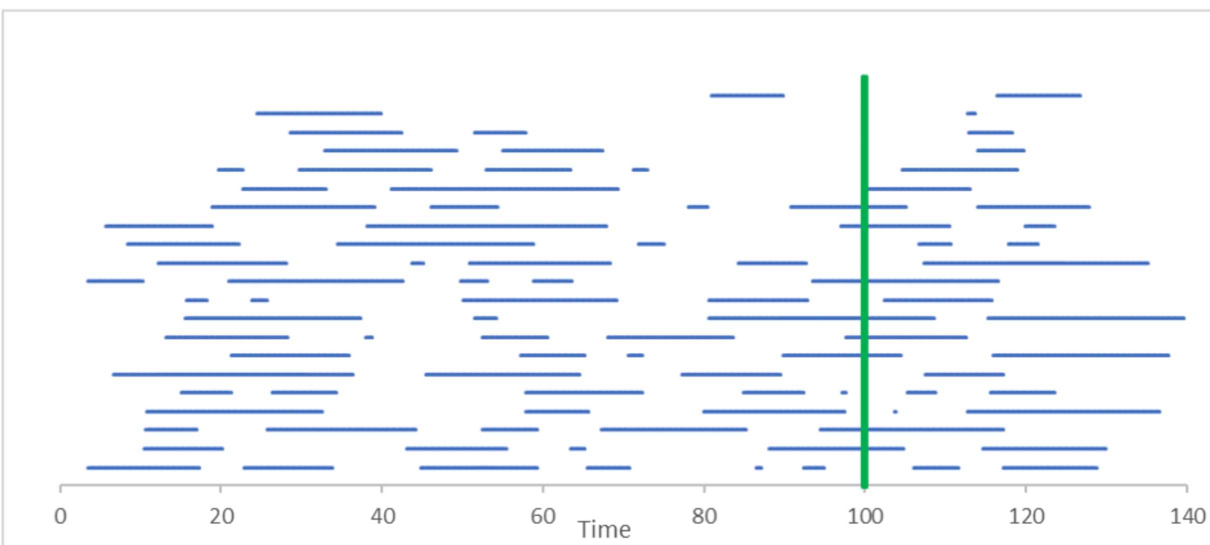

Figure 4: Example of 100 event durations typical of Set 8, distributed as U[0, N[20, 5]], constant population start times uniformly distributed in a timespan from 0 to 120. Sample time is at 100.

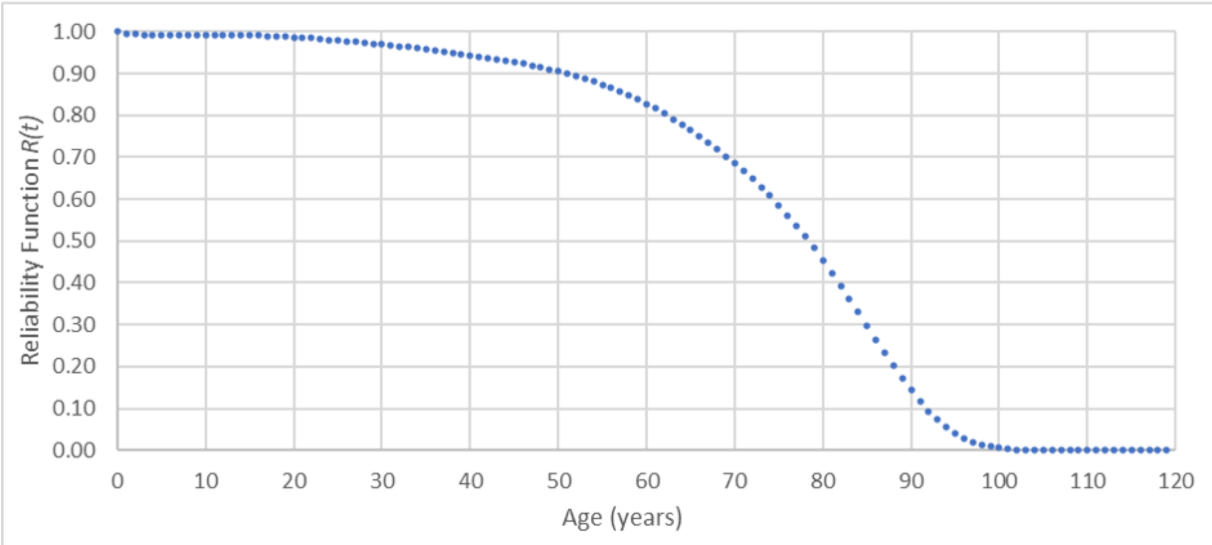

Figure 5: Reliability function $R(t)$ for male lifetimes in the United States based on Social Security Administration data.

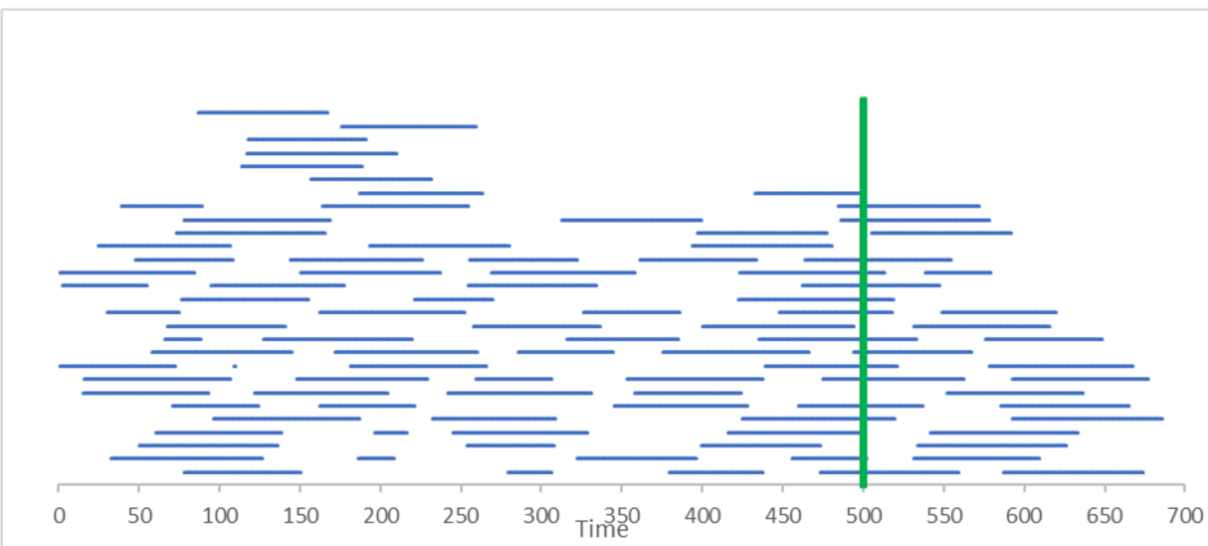

Figure 6: Example of 100 event durations typical of Set 9, generated using the reliability function for male lifetimes in the United States shown in Figure 5, with constant population start times uniformly distributed in a simulation calendar timespan from 0 to 600. Sample is taken at simulation calendar time 500.

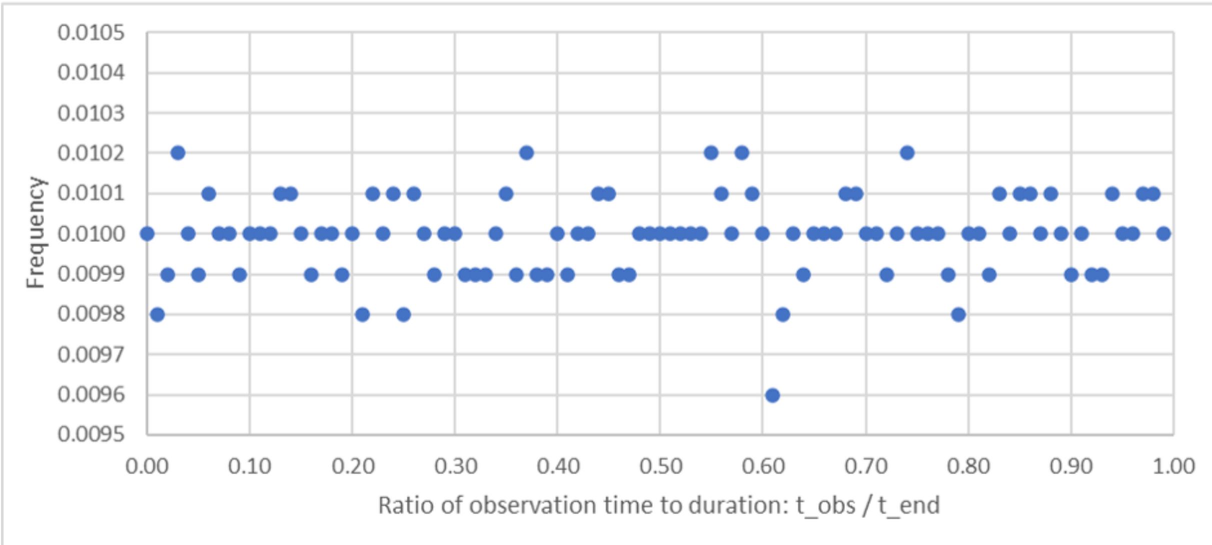

Figure 7: Histogram of the frequencies of $\frac{t_{obs}}{t_{end}}$ for 100 bins in the interval [0, 1], based on 1,000,000 Monte Carlo simulations, for Set 1 event durations distributed as U[0, 20], constant population.

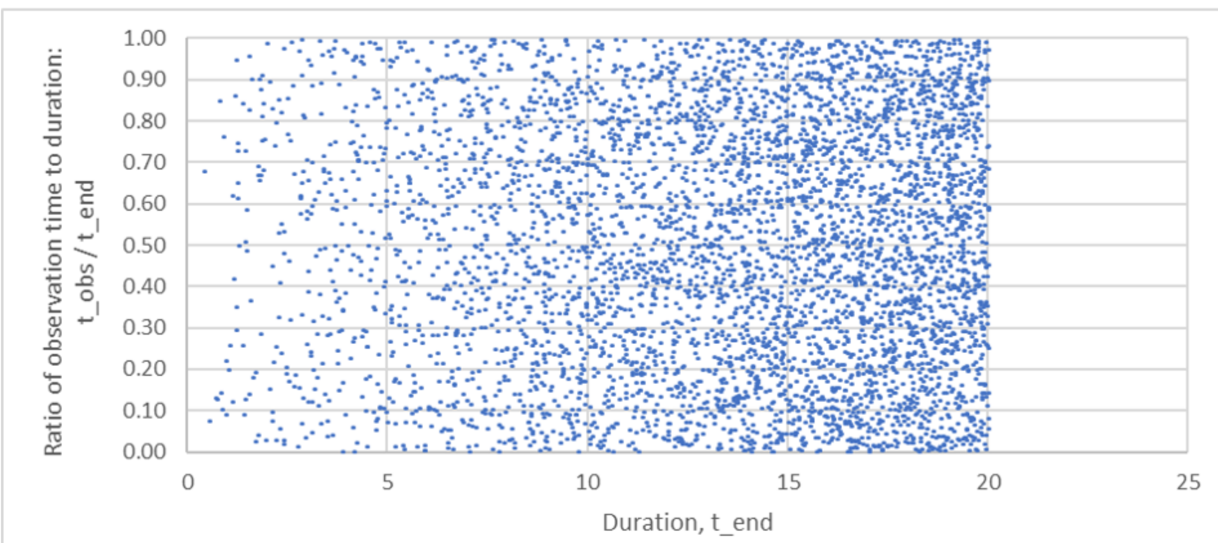

Figure 8: Scatter plot of $\frac{t_{obs}}{t_{end}}$ as a function of t_{end}, based on 5,000 Monte Carlo simulations, for Set 1 event durations distributed as U[0, 20], constant population.

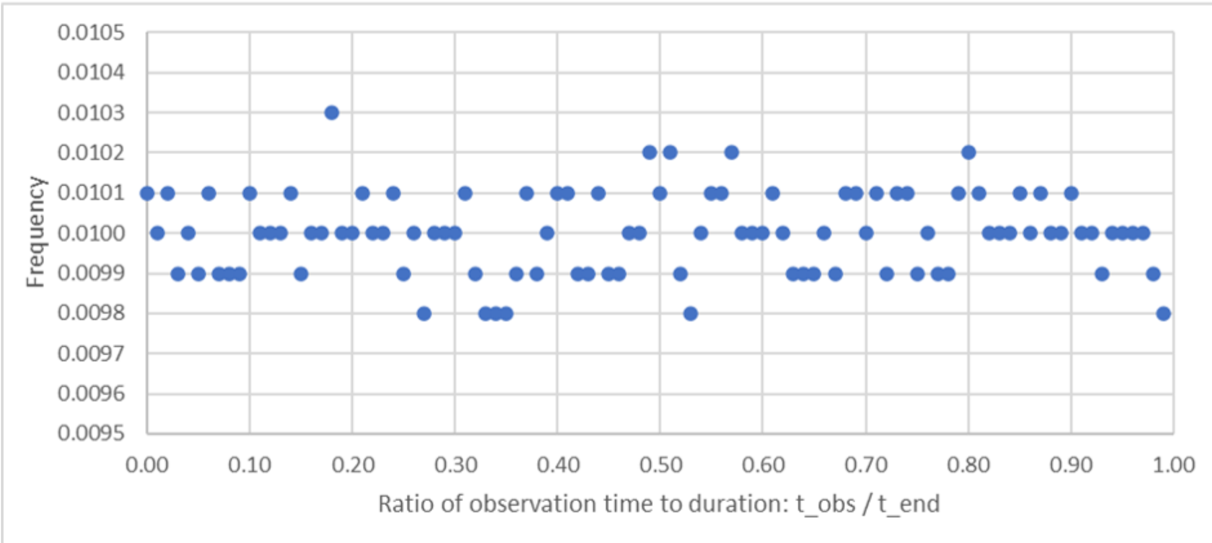

Figure 9: Histogram of the frequencies of $\frac{t_{obs}}{t_{end}}$ for 100 bins in the interval [0, 1], based on 1,000,000 Monte Carlo simulations, for Set 2 event durations exponentially distributed with MTTF = $1/\lambda = 10$, constant population.

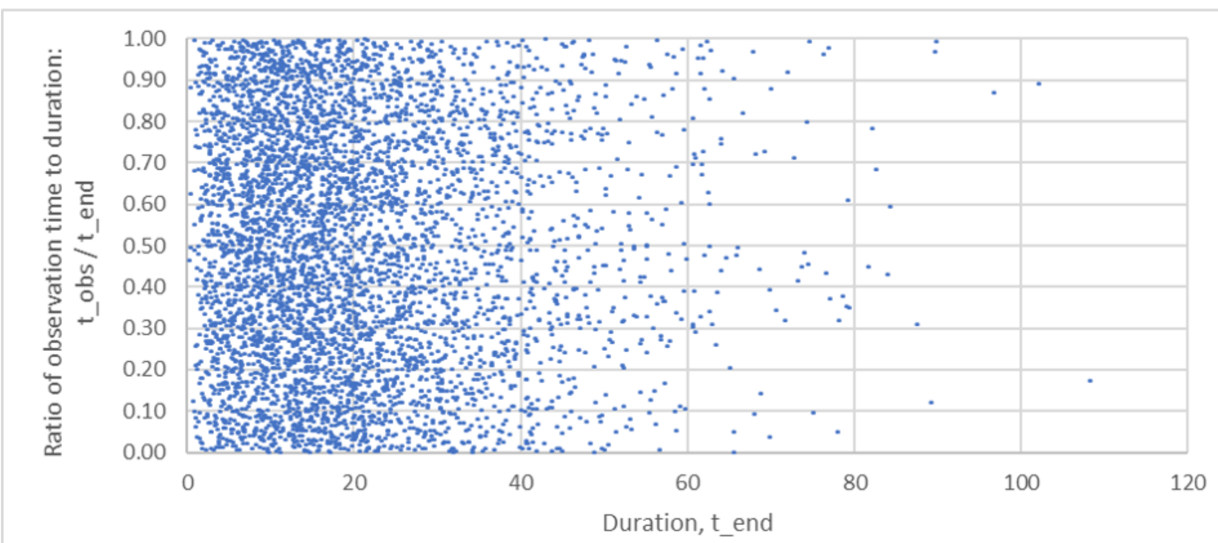

Figure 10: Scatter plot of $\frac{t_{obs}}{t_{end}}$ as a function of t_{end}, based on 5,000 Monte Carlo simulations, for Set 2 event durations exponentially distributed with MTTF = $1/\lambda = 10$, constant population.

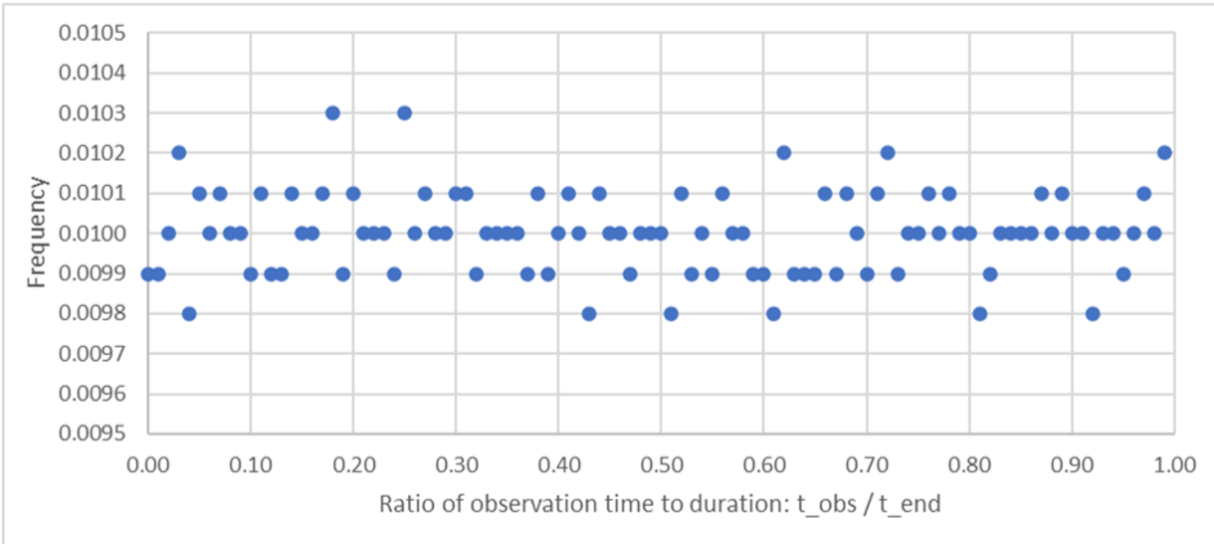

Figure 11: Histogram of the frequencies of $\frac{t_{obs}}{t_{end}}$ for 100 bins in the interval [0, 1], based on 1,000,000 Monte Carlo simulations, for Set 4 event durations distributed as N[10, 2], constant population.

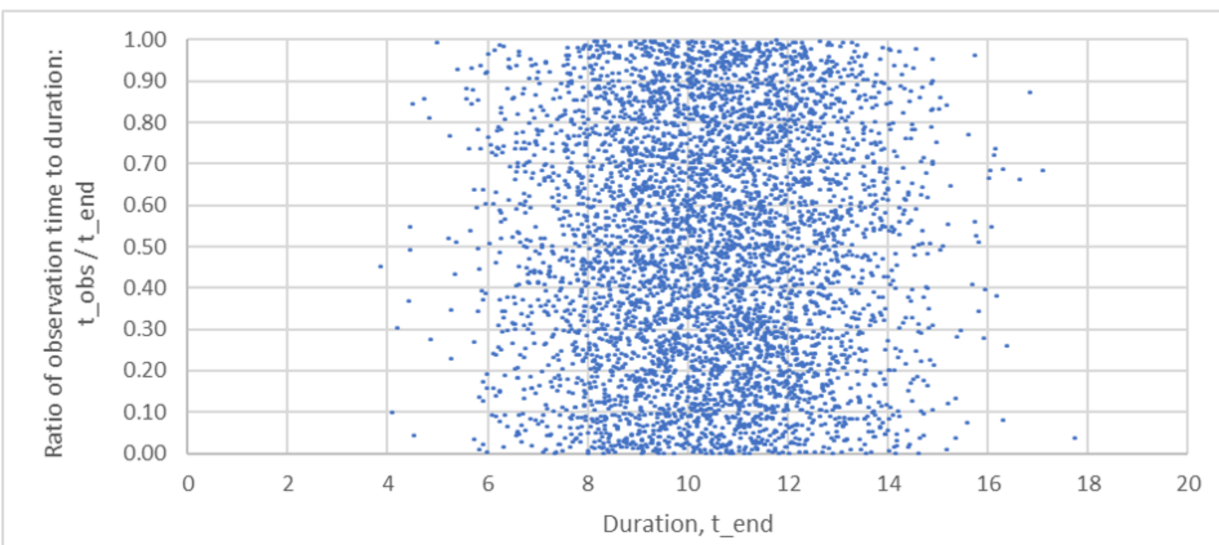

Figure 12: Scatter plot of $\frac{t_{obs}}{t_{end}}$ as a function of t_{end}, based on 5,000 Monte Carlo simulations, for Set 4 event durations distributed as N[10, 2], constant population.

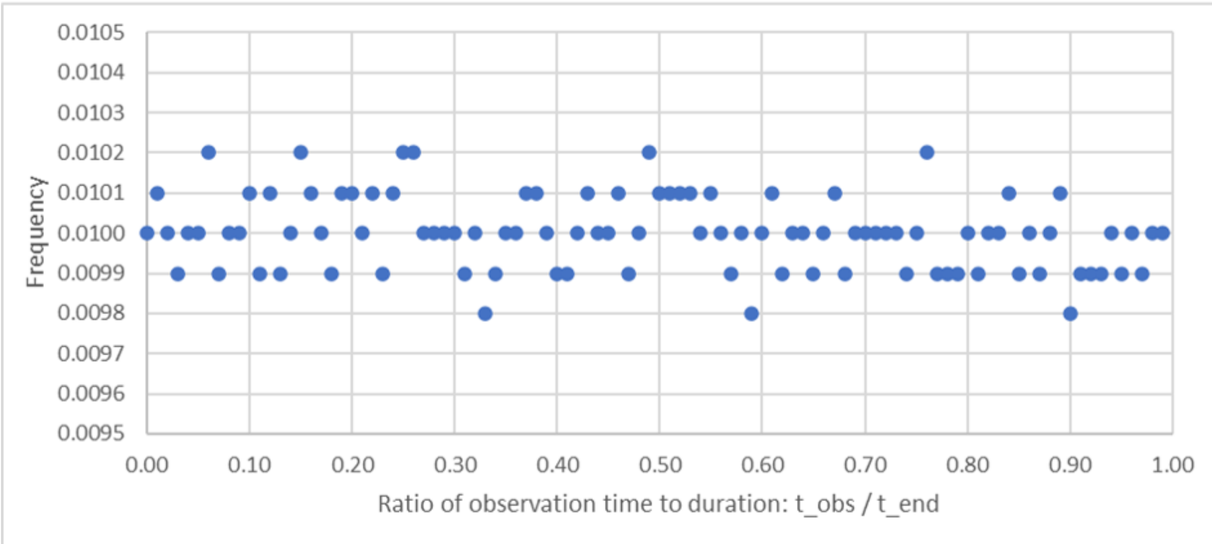

Figure 13: Histogram of the frequencies of $\frac{t_{obs}}{t_{end}}$ for 100 bins in the interval [0, 1], based on 1,000,000 Monte Carlo simulations, for Set 8 event durations distributed as U[0, N[20, 5]], constant population.

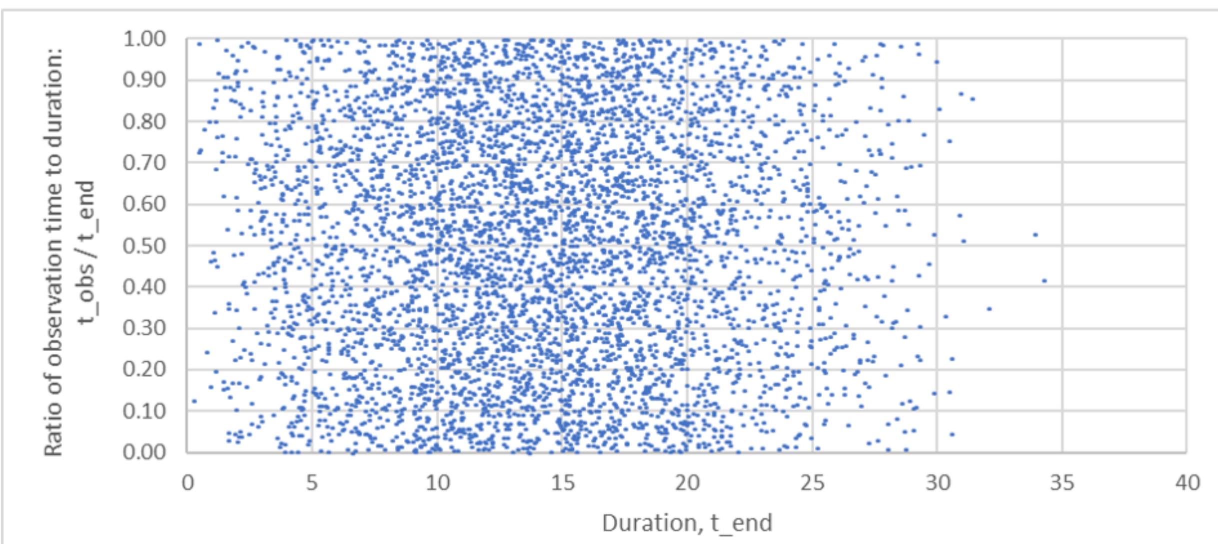

Figure 14: Scatter plot of $\frac{t_{obs}}{t_{end}}$ as a function of t_{end}, based on 5,000 Monte Carlo simulations, for Set 8 event durations distributed as U[0, N[20, 5]], constant population.

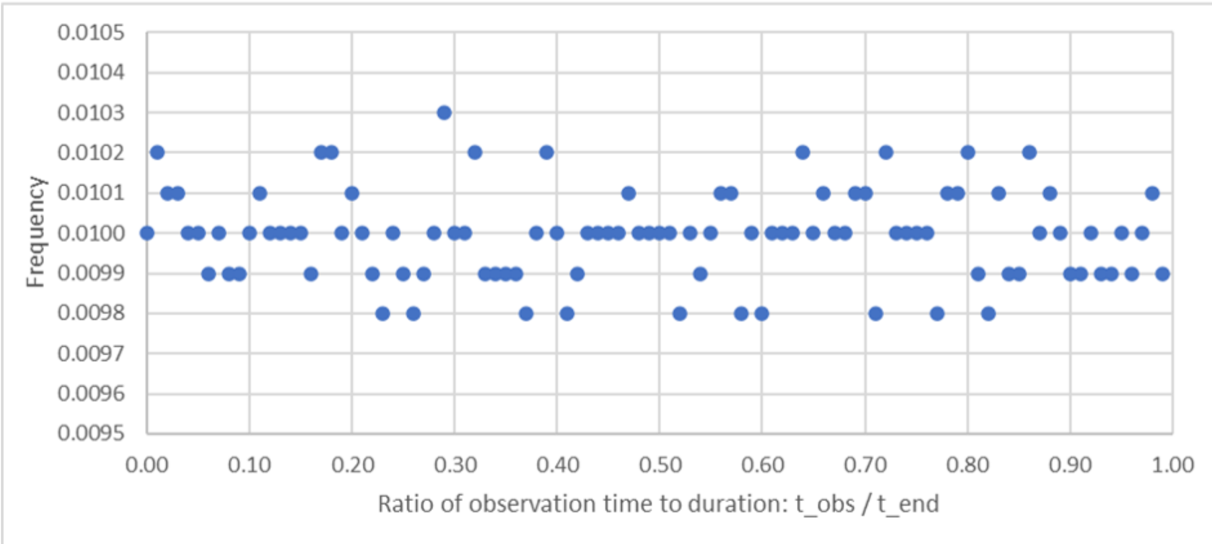

Figure 15: Histogram of the frequencies of $\frac{t_{obs}}{t_{end}}$ for 100 bins in the interval [0, 1], based on 1,000,000 Monte Carlo simulations, for Set 9 event durations generated using the reliability function for male lifetimes in the United States shown in Figure 5, constant population.

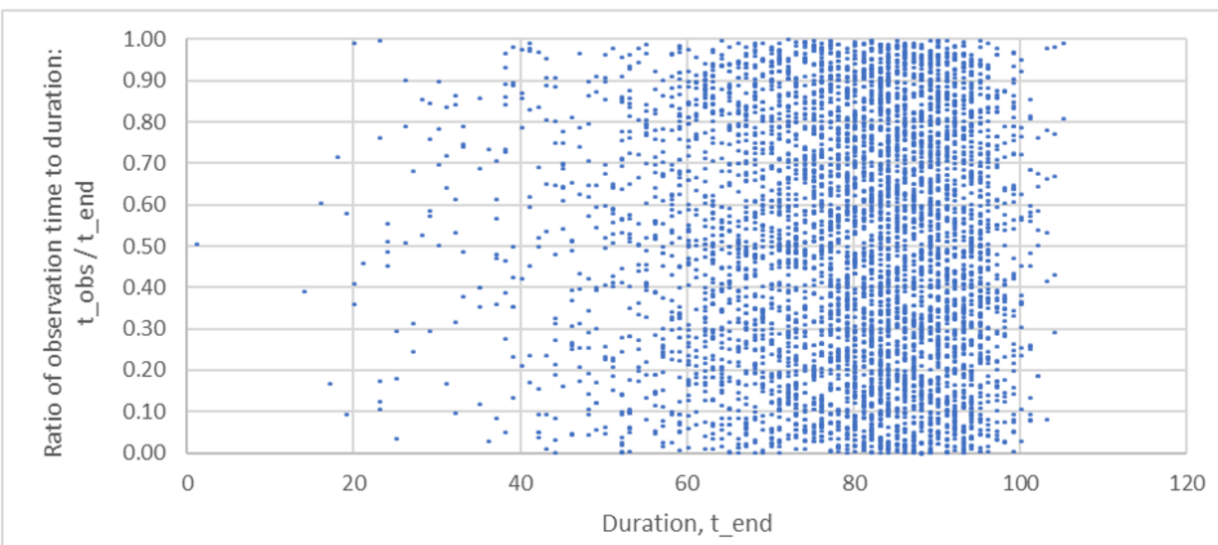

Figure 16: Scatter plot of $\frac{t_{obs}}{t_{end}}$ as a function of t_{end}, based on 5,000 Monte Carlo simulations, for Set 9 event durations generated using the reliability function for male lifetimes in the United States shown in Figure 5, constant population.

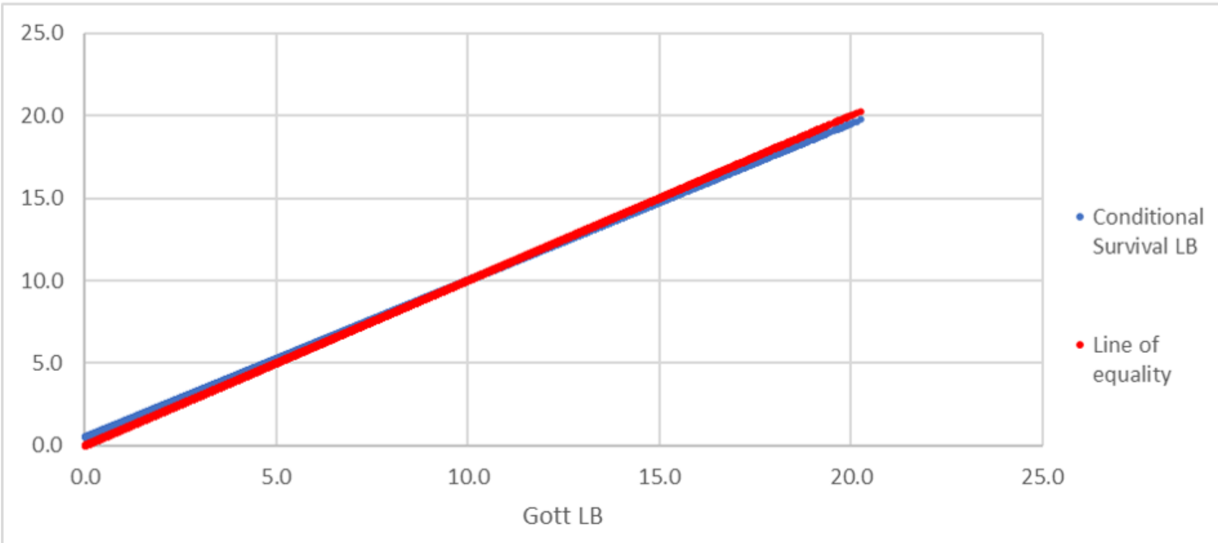

Figure 17: Scatter plot of conditional survival LBs as a function of Gott LBs, based on 10,000 Monte Carlo simulations, for Set 1 event durations distributed as U[0, 20], constant population. Red line would be equality with Gott's LBs.

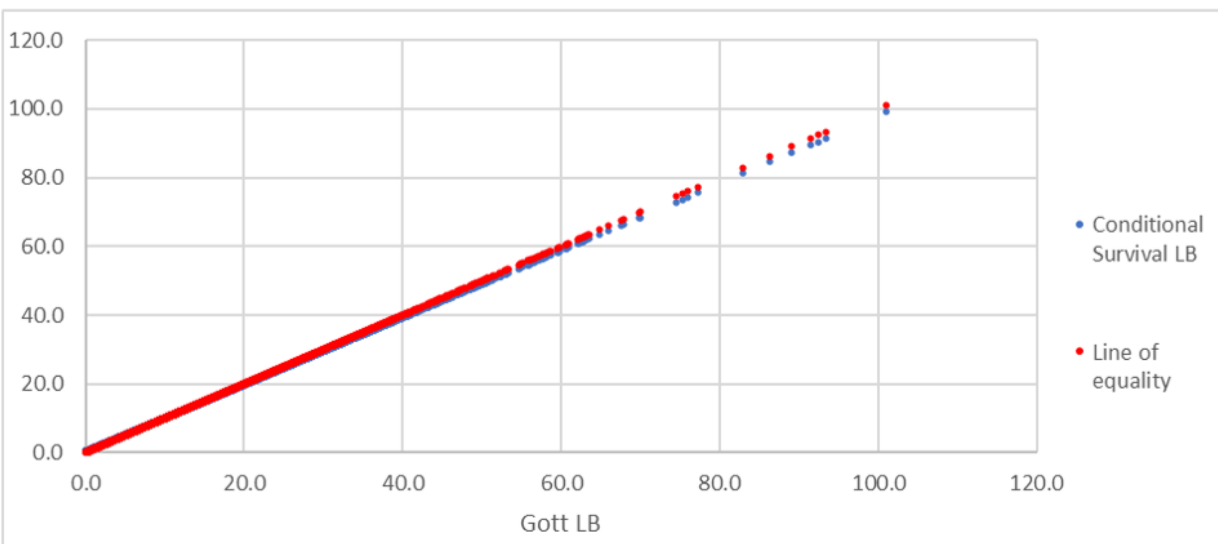

Figure 18: Scatter plot of conditional survival LBs as a function of Gott LBs, based on 10,000 Monte Carlo simulations, for Set 2 event durations exponentially distributed with MTTF = $1/\lambda = 10$, constant population. Red line would be equality with Gott's LBs.

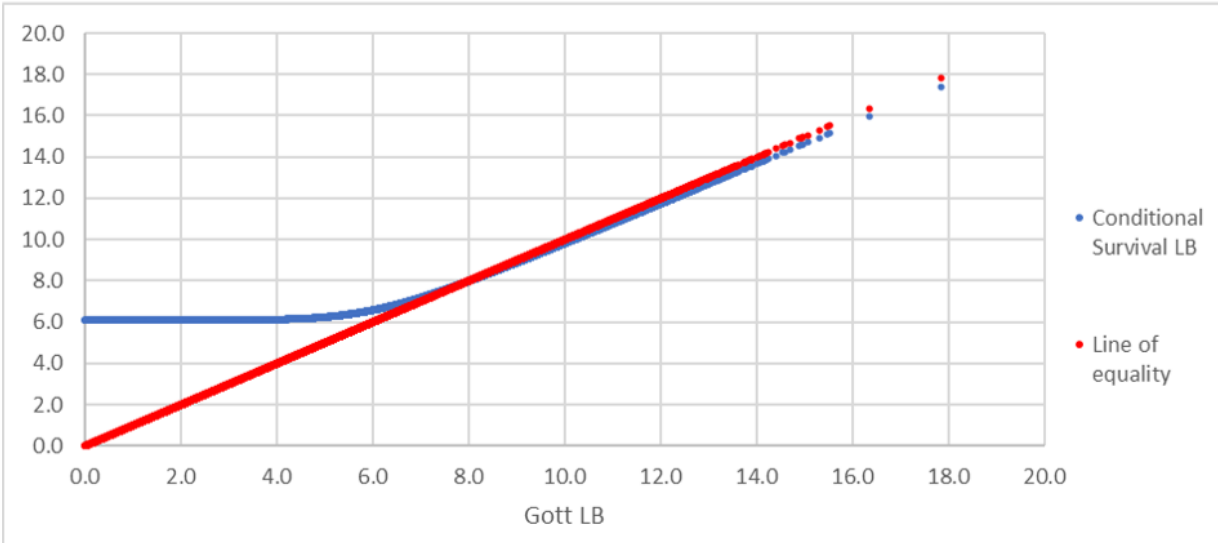

Figure 19: Scatter plot of conditional survival LBs as a function of Gott LBs, based on 10,000 Monte Carlo simulations, for Set 4 event durations distributed as N[10, 2], constant population. Red line would be equality with Gott's LBs.

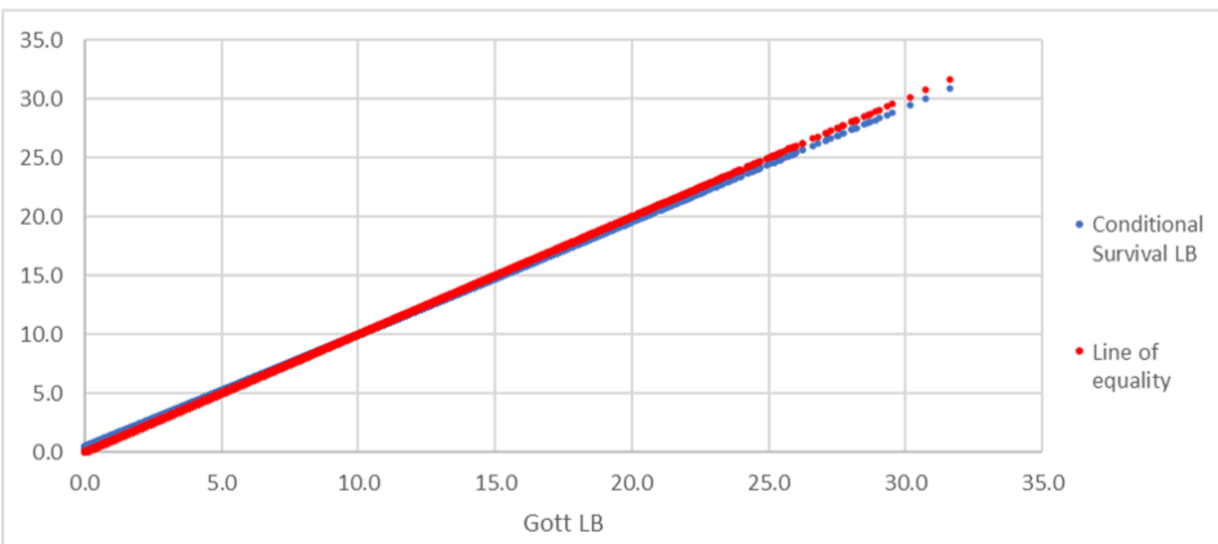

Figure 20: Scatter plot of conditional survival LBs as a function of Gott LBs, based on 10,000 Monte Carlo simulations, for Set 8 event durations distributed as U[0, N[20, 5]], constant population. Red line would be equality with Gott's LBs.

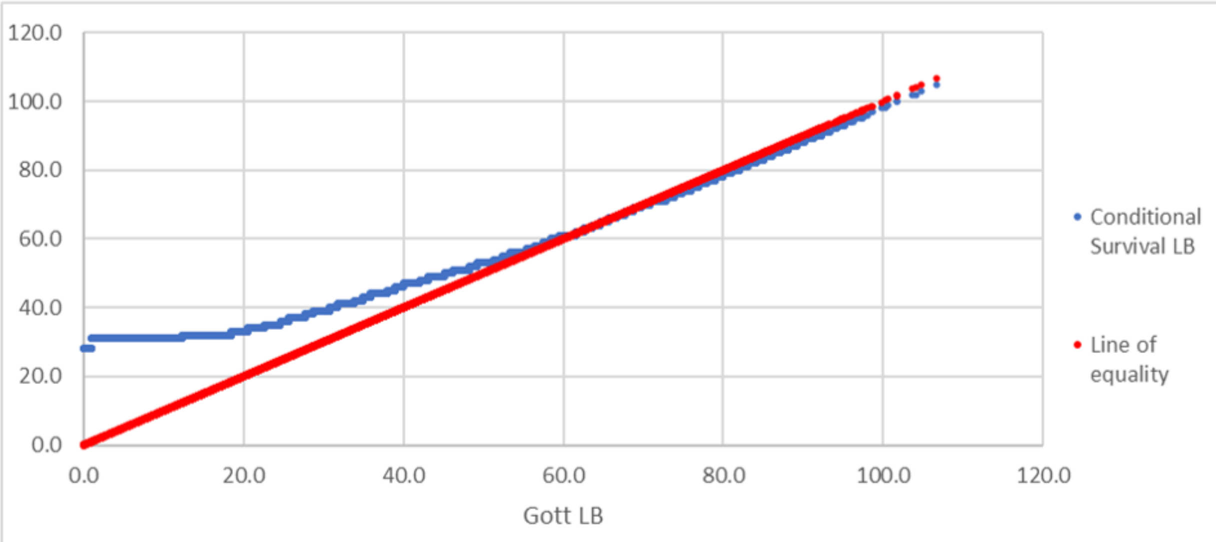

Figure 21: Scatter plot of conditional survival LBs as a function of Gott LBs, based on 10,000 Monte Carlo simulations, for Set 9 event durations generated using the reliability function for male lifetimes in the United States shown in Figure 5, constant population. Red line would be equality with Gott's LBs.

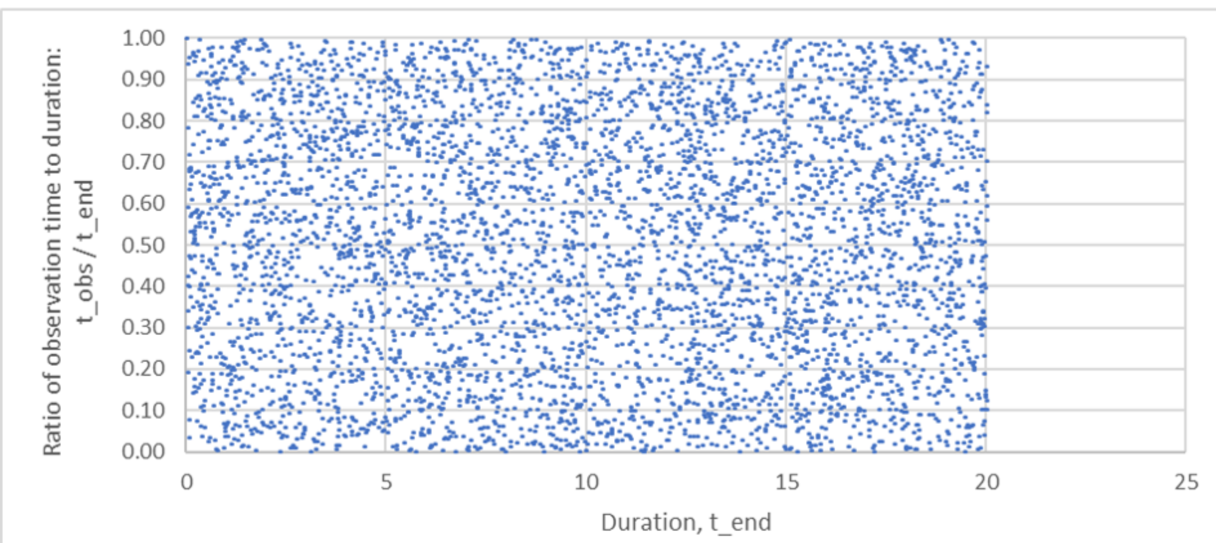

Figure 22: Scatter plot of $\frac{t_{obs}}{t_{end}}$ as a function of t_{end}, based on 5,000 Monte Carlo simulations, for Set 1 event durations distributed as U[0, 20], for a constant population using the Copernican Observation model. Unlike Figure 8 for the Copernican Sample model, this figure shows apparently uniform density of points across the Duration axis as well as across the Ratio axis.

7 Populations of Events with Baby Booms

Until this point, it has been assumed that Gott's method strictly requires that the sample time of an event occur at no special time. Section 6 showed that sampling an event from a constant population with uniformly distributed start times satisfies the Copernican Principle.

Now the problem is considered in cases where the Copernican Principle is violated to some extent to determine if Gott's method still works.

This section shows that Gott's method has a degree of tolerance against the circumstance of sampling from a population that is not constant or in a steady state. It is shown that Gott's method can be tolerant of a bias toward observing an event somewhat earlier or later in its duration.

A dictionary definition of a *baby boom* is "a marked rise in birthrate" [Merriam-Webster (2005)]. This paper applies two "baby boom" models for event duration start times.

The first baby boom model uses start times that are *normally distributed* as N[μ_{BB}, σ_{BB}] for event durations governed by several distributions.

A second baby boom model uses start times from a distribution with peaks and troughs based on historical *human birth data* in the United States; these are merged data for annual numbers of births provided by the Centers for Disease Control (CDC), Infoplease, and Statistica [CDC (2023)] and are shown in Figure 52.[30] These highly irregularly distributed start times are used to test Gott's method with the previously discussed SSA data [SSA (2023)] for the human lifetime distribution shown in Figure 5.

The baby boom populations simulated for this work correspond to Sets 10 through 25 defined in Table 1.

[30] This is not to say that the recent SSA data used for the event duration reliability function would be stable over the timespan of the CDC et al. birth data. Furthermore, the birth data combine all births in the United States in those years. The purpose of this exercise is only to show a non-trivial distribution for event durations with a non-trivial distribution for baby boom start times. The data are fairly realistic, but should be considered only notional.

In Section 6 for constant populations, Figure 1 showed 100 event durations typical of Set 1, uniformly distributed as U[0, 20] where the start times were also uniformly distributed in a simulation calendar timespan from 0 to 120 with a fixed sample time at 100.

In this section, however, Figure 23 shows the same situation as in Figure 1 but with event durations typical of Set 11, where start times are drawn instead from N[80, 20], meaning they are normally distributed with a standard deviation of 20 and their mean at 80 is centered 1 standard deviation before the sample time. This creates a "baby boom" of events with the peak of its start times at 1 standard deviation *before* the sample is taken. Rather than sampling uniformly across event durations, this sample time causes the observer to tend to capture "older" events, meaning later in their durations, than a similar sample taken from a constant population.

Figure 24 shows the same set of durations, but with the mean of their start times centered *at* the sample time at 100, so they are typical of Set 12.

Figure 25 again shows the same set of durations, but with the peak of their start times at 1 standard deviation *after* the sample time at 100, so they are typical of Set 13. This sample time causes the observer to tend to sample "younger" events, meaning earlier in their durations, than a similar sample taken from a constant population. (The situation shown in Figure 24 also tended to sample "younger" events.)

Other than being shifted in time relative to the sample time, Figure 23, Figure 24, and Figure 25 are identical to each other because the same random seed was used to generate them. This random seed was different from that used to generate Figure 1.

The same approach is used to produce Figure 26, Figure 27, and Figure 28 for 100 events typical of Sets 16, 17, and 18, with durations exponentially distributed with MTTF $= 1/\lambda = 10$. This approach is used again in Figure 29, Figure 30, and Figure 31 to show 100 events with durations normally distributed as N[10, 5].

The plots in Figure 23 through Figure 31 are provided only to help visualize how the Monte Carlo simulations were implemented. They are similar to Figure 1, Figure 2, Figure 3, Figure 4, and Figure 6 used in Section 6.

The next set of plots give the results of the simulations.

In each case for the plots in Figure 32 through Figure 51, the population start times are normally distributed with $\sigma_{BB} = 20$ and the sample time is 100. The plots are in groups of either five figures or five pairs of figures, where the values of μ_{BB} are 60, 80, 100, 120, and 140 so that the peaks of the baby booms are at 2 and 1 standard deviation *before* the sample time, directly *at* the sample time, and 1 and 2 standard deviations *after* the sample time.

Although the plotted examples of 100 event durations used the same random seeds for each distribution to show how the populations shifted with the baby booms, different random seeds were used for each set of Monte Carlo simulations.

The five pairs of plots in Figure 32 through Figure 41 correspond to the five Sets 10 through 14. These sets are for event durations uniformly distributed as U[0, 20], as was the case for Set 1, but with baby boom start times instead of uniformly distributed start times.

As described in Section 6, the first figure in each pair is a crude histogram of the frequencies of the ratios $\frac{t_{obs}}{t_{end}}$ for 100 bins in the interval [0, 1] for the 1,000,000 Monte Carlo simulations performed for each set. The second in each pair is a scatter plot of $\frac{t_{obs}}{t_{end}}$ as a function of t_{end} for the first 5,000 of the Monte Carlo simulations performed for each set.

The 100 frequency bins in the histograms shown in Section 6 for constant populations hovered closely about the value of 0.01, which would be necessary for $\frac{t_{obs}}{t_{end}}$ to be uniformly distributed on the interval [0, 1], but the frequency bins in these histograms for populations with baby booms do not share that property: the 100 frequencies for Sets 10 through 14 are generally centered around 0.01 but their trends visibly deviate considerably from being flat enough to satisfy the Copernican Principle.[31]

The scatter plots for Sets 10 through 14 also differ from those shown in Section 6 in that the densities of points on the Ratio axes are not uniform along those axes. This emphasizes the fact that t_{obs} does not

[31] Although probably unneeded in these cases, the χ^2 test for the bins obtained for each of Sets 10 through 14 yields a p-value of 0, and thus one can reject the null hypothesis that the deviation from the expected values is due to chance alone; these ratios for baby boom populations are not uniformly distributed.

occur uniformly within the interval [0, t_{end}] when the start times are non-uniformly distributed.

Similar deviations from uniform distribution of the ratios $\frac{t_{obs}}{t_{end}}$ are demonstrated for Sets 15 through 19 having event durations exponentially distributed with MTTF = $1/\lambda$ = 10, and baby boom start times, in Figure 42 through Figure 46.[32]

These deviations from uniform distribution of $\frac{t_{obs}}{t_{end}}$ are also demonstrated for Sets 20 through 24 having event durations normally distributed as N[10, 2], and baby boom start times, in Figure 47 through Figure 51.

These examples of distributions with baby boom start times violate the only necessary assumption of Gott's method. The consequences of this are shown in Table 3 which gives the actual 2.5%ile LB and 97.5%ile UB on ρ for Gott's method based on the Monte Carlo simulations for the uniform, exponential, and truncated normal distributions tested in baby boom Sets 10 through 24. These rows show that both the actual LB and UB for ρ are smaller than the expected ρ_{LB} and ρ_{UB} of the CI (4) for the "older" events sampled from an earlier baby boom peak and larger than the expected ρ_{LB} and ρ_{UB} for the "younger" events sampled from a later baby boom peak.

An unexpected result is shown, however, for these same baby boom Sets 10 through 24 in Table 4. This table gives the confidence levels and tail probabilities resulting from using Gott's CIs given as (4). The confidence levels achieved by Gott's ρ, shown in the second column, are still extremely close to the target of 95% *in spite of* these violations of the Copernican Principle.

The tail probabilities given in the third and fourth columns, on the other hand, do not achieve the target values of 2.5% although they are not very far from those target values. The lower tail probability (the probability that the actual LB on ρ is lower than the target LB) is larger than 2.5% for early baby boom peaks and is smaller than 2.5% for later baby boom peaks; the trend is reversed for the upper tail probabilities. The CIs based on (4) for the non-uniformly distributed baby boom start times may still achieve close to their target confidence level of 95% but they are now slightly non-central.

[32] Only the histograms are shown; the scatter plots are omitted as the histograms are sufficient to show the deviations from uniform distribution.

The results for baby boom Sets 10 through 24 involve simple distributions for event durations and normally distributed start times for baby booms. To model a more complex situation, this phenomenon was explored in Set 25 using the $R(t)$ for male lifetimes in the United States based on the SSA data [SSA (2023)] shown in Figure 5 and employed in Section 6 for constant populations. In this section, however, the start times for those event durations are generated using the CDC et al. birth data depicted in Figure 52 which include several historical human baby booms.

For Set 25, Table 4 gives the confidence level and tail probabilities for these SSA-based human lifetime event durations with the CDC et al. birth data. The confidence level in the second column is 0.942, which is extremely close to the target value of 95%. The lower tail probability of 0.026 in the third column is very close to the target of 2.5%, while the upper tail probability of 0.032 in the fourth column is slightly larger than the target of 2.5%.

Thus, the results of the Monte Carlo simulations for baby boom Sets 10 through 25 showed that the CIs still provide close to the target level of 95% but they have also been shown to be non-central as their tail probabilities diverge slightly from the target of 2.5%.

Non-central CIs with unpredictable tail probabilities are not as useful as central CIs, or even non-central CIs with known tail probabilities. However, these demonstrations have shown that, even when the Copernican Principle is violated to some extent *in addition to* the underlying distribution or reliability function for the event durations being unknown, Gott's method can still provide CIs that are robust enough to give valid 95% CIs. The cost in these cases seems to be that one can make only the weaker statement that the remaining duration is within the LB and UB, but cannot state with close accuracy the probability of being below the LB or above the UB.

When the underlying distribution is known, another method such as conditional survival would be the appropriate tool. Table 4 shows for the baby boom Sets 10 through 25 that the confidence levels in the fifth column and the tail probabilities in the sixth and seventh column are unaffected by the non-uniformly distributed start times, as would be expected. They achieve the target values, with only Set 25 having a slightly lower confidence of 0.949 and tail probabilities that are slightly non-central.

The sampling of event durations in populations with baby boom start times resulted in different statistics from those observed for the constant populations. For example, Set 1, used in Section 6, had event durations uniformly distributed as U[0, 20] in a constant population with uniformly distributed start times; the average value of t_{obs} was 6.70 and the average value of $\frac{t_{obs}}{t_{end}}$ was 0.50. As noted previously in this section, the baby boom Sets 10 through 14 had event durations with the same distribution as in Set 1, but had baby boom start times that peaked at 60, 80, 100, 120, and 140.[33] The average values of t_{obs} for these baby boom sets were 8.51, 7.35, 6.23, 5.34, and 4.51, and the average values of $\frac{t_{obs}}{t_{end}}$ were 0.59, 0.53, 0.48, 0.43, and 0.39.

This example emphasizes again that an early baby boom results in "older" events being observed and a later baby boom results in "younger" events being observed. And yet, in spite of the violation of the Copernican Principle by the non-uniformly distributed start times, Table 4 shows that the confidence levels achieved by Gott's CIs given as (4) are still very close to the target of 95%.

[33] With the sample time at 100, the peaks of the baby booms are at 2 and 1 standard deviation before the sample time, directly at the sample time, and 1 and 2 standard deviations after the sample time.

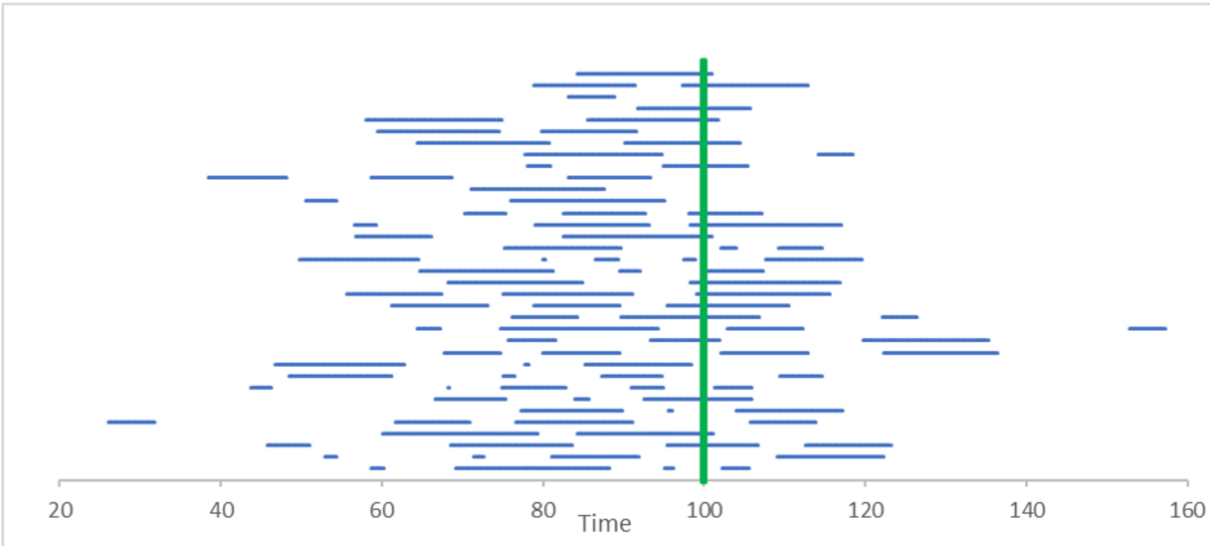

Figure 23: Example of 100 event durations typical of Set 11, distributed as U[0, 20], baby boom start times distributed as N[80, 20]. Mean of start times is 1 SD before sample time at 100.

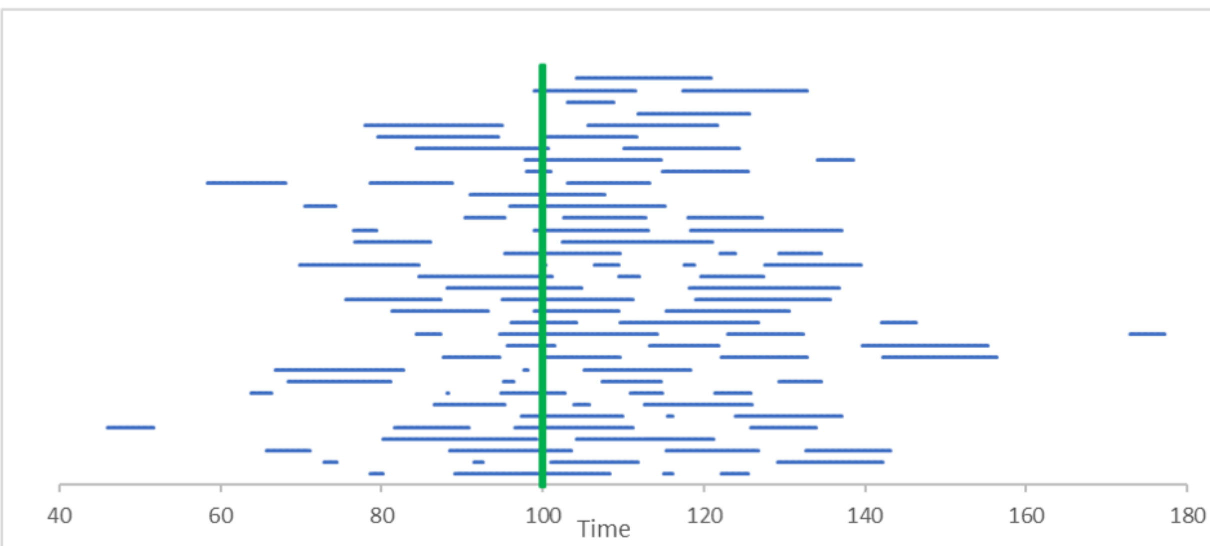

Figure 24: Example of 100 event durations typical of Set 12, distributed as U[0, 20], baby boom start times distributed as N[100, 20]. Mean of start times is at sample time at 100.

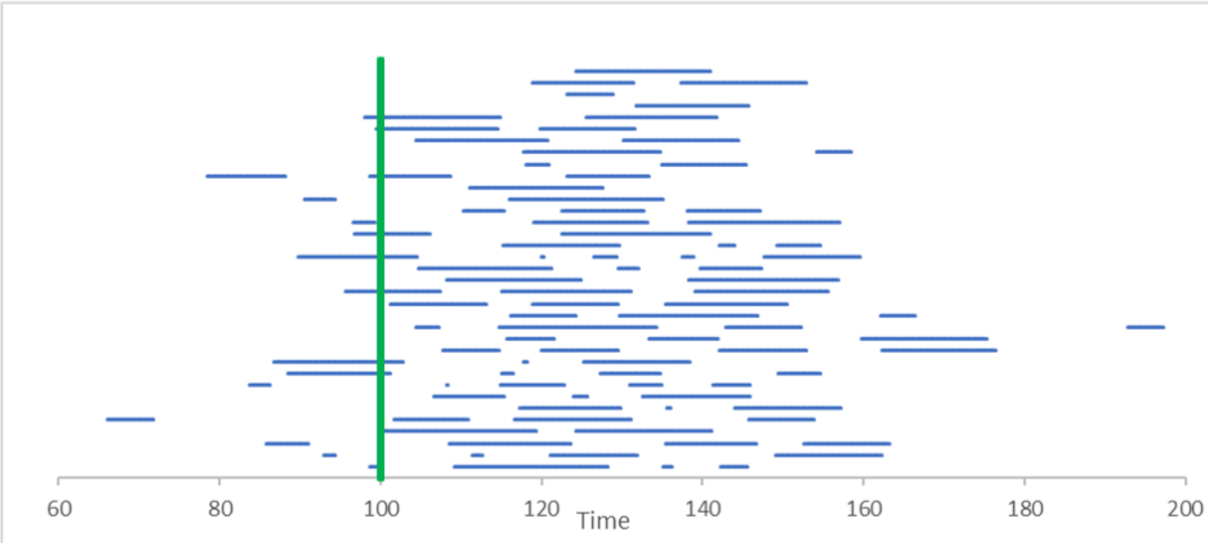

Figure 25: Example of 100 event durations typical of Set 13, distributed as U[0, 20], baby boom start times distributed as N[120, 20]. Mean of start times is 1 SD after sample time at 100.

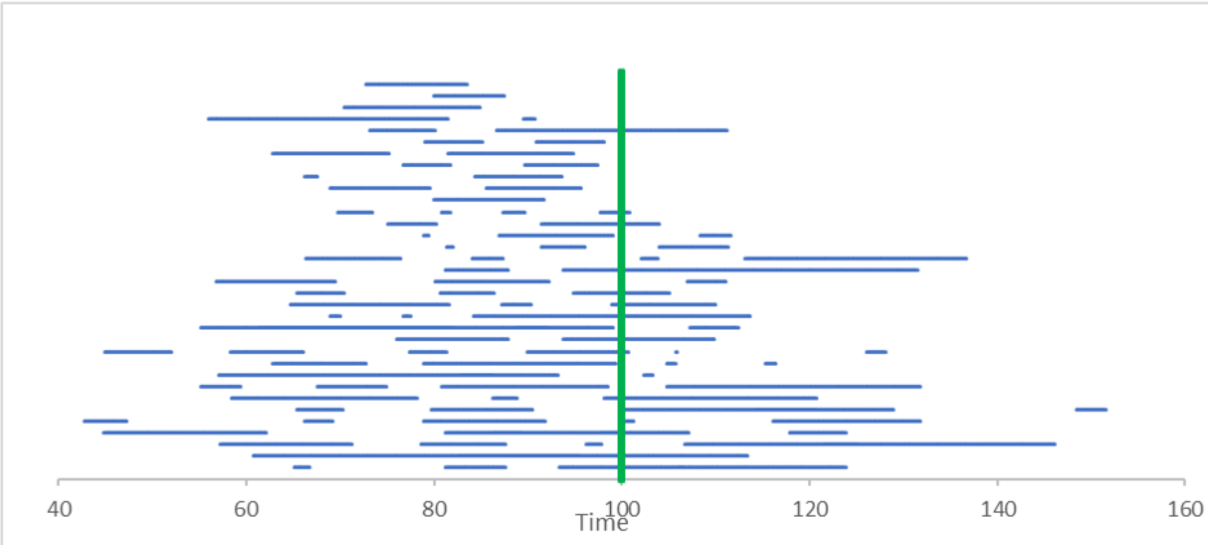

Figure 26: Example of 100 event durations typical of Set 16, exponentially distributed with MTTF = $1/\lambda = 10$, baby boom start times distributed as N[80, 20]. Mean of start times is 1 SD before sample time at 100.

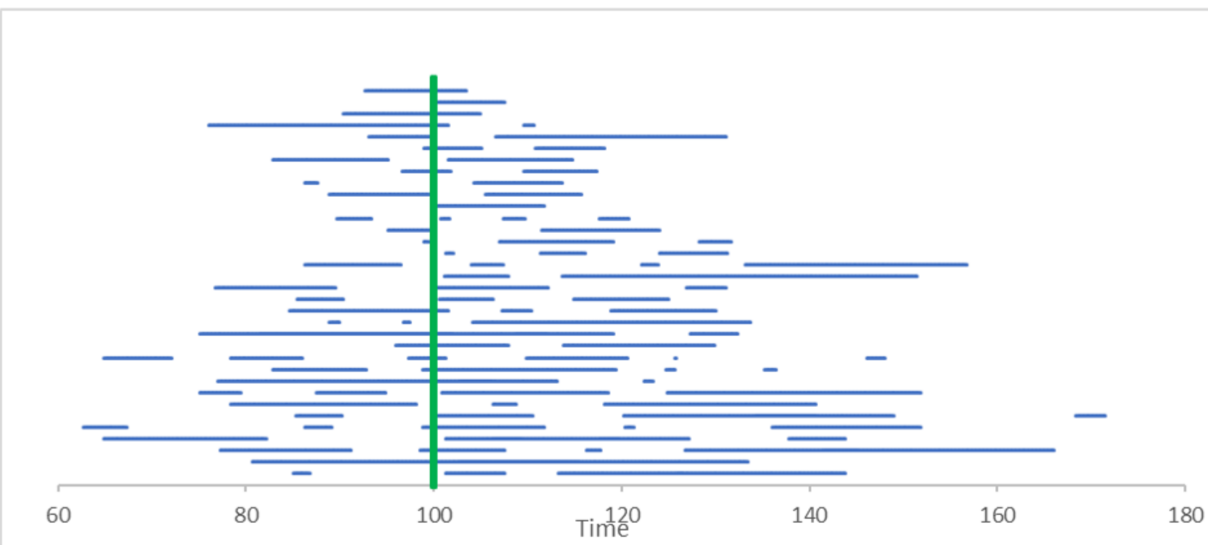

Figure 27: Example of 100 event durations typical of Set 17, exponentially distributed with MTTF = $1/\lambda = 10$, baby boom start times distributed as N[100, 20]. Mean of start times is at sample time at 100.

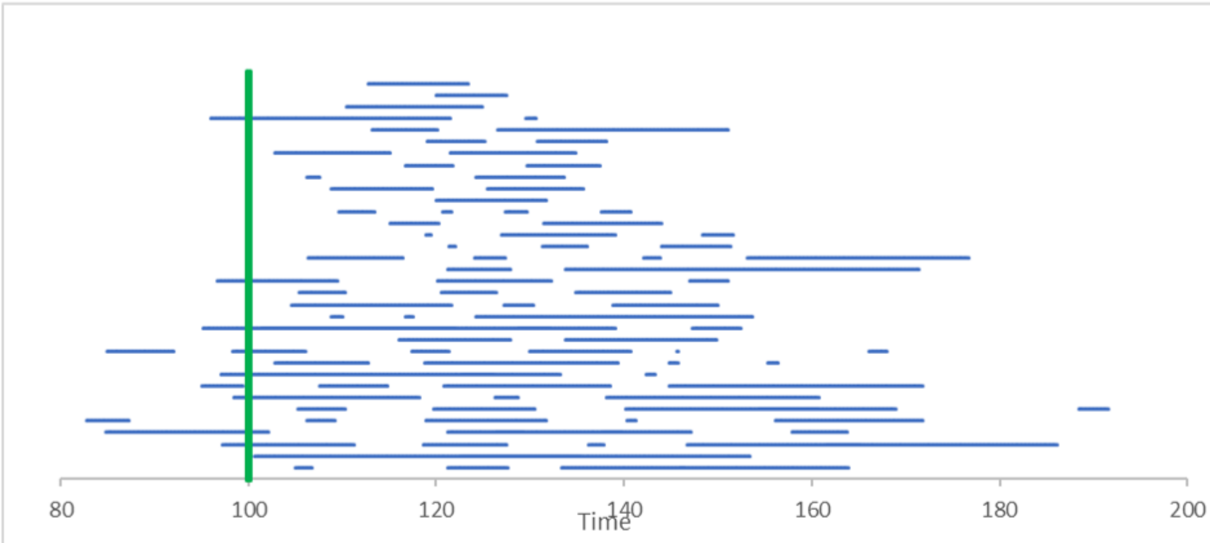

Figure 28: Example of 100 event durations typical of Set 18, exponentially distributed with MTTF = $1/\lambda = 10$, baby boom start times distributed as N[120, 20]. Mean of start times is 1 SD after sample time at 100.

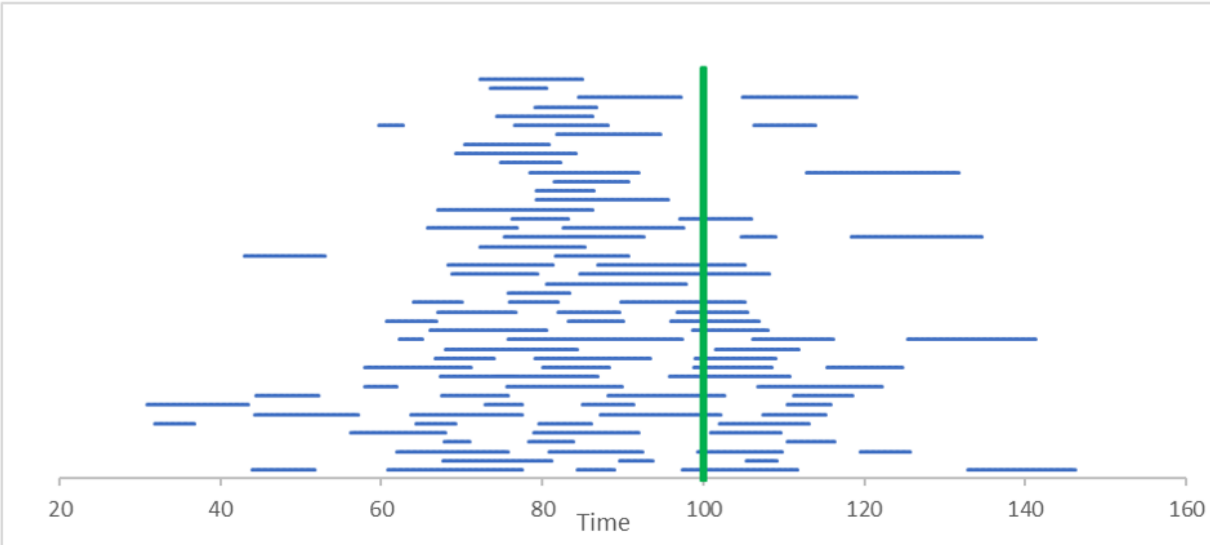

Figure 29: Example of 100 event durations distributed as N[10, 5], baby boom start times distributed as N[80, 20]. Mean of start times is 1 SD before sample time at 100.

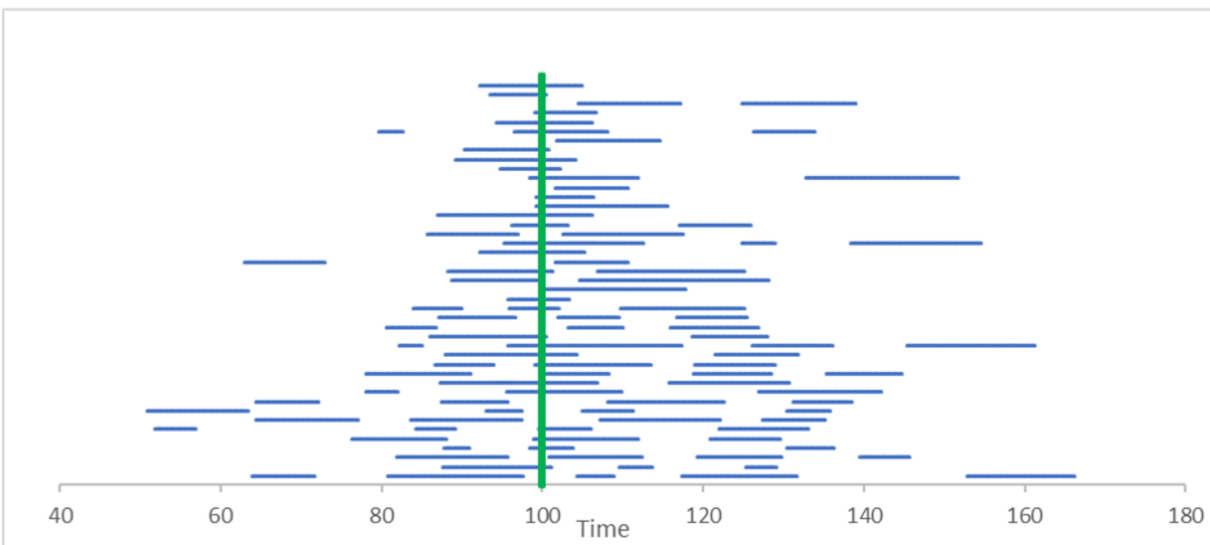

Figure 30: Example of 100 event durations distributed as N[10, 5], baby boom start times distributed as N[100, 20]. Mean of start times is at sample time at 100.

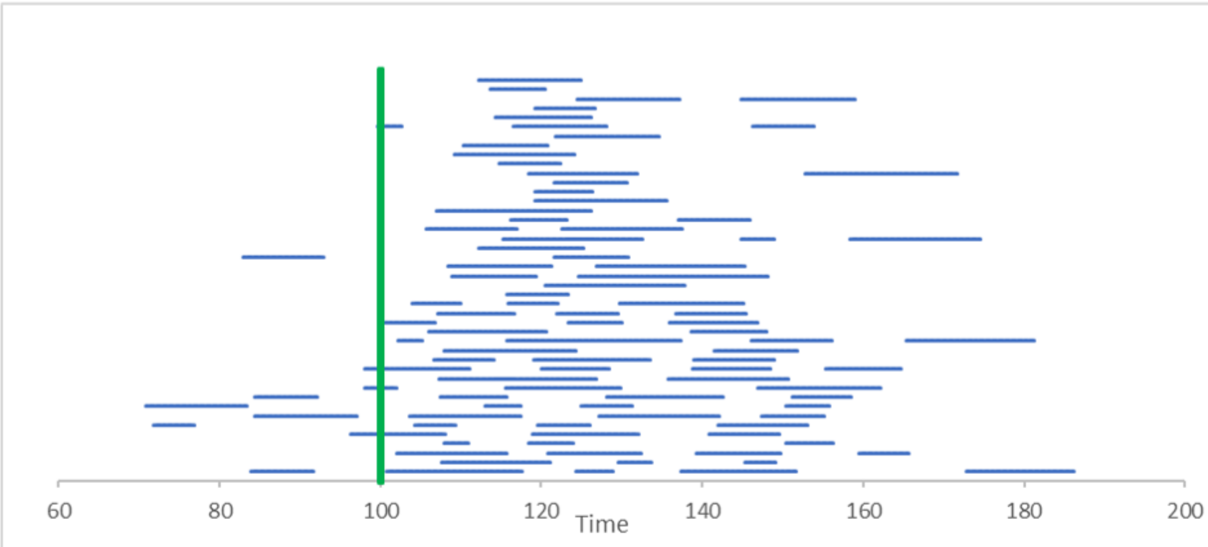

Figure 31: Example of 100 event durations distributed as N[10, 5], baby boom start times distributed as N[120, 20]. Mean of start times is 1 SD after sample time at 100.

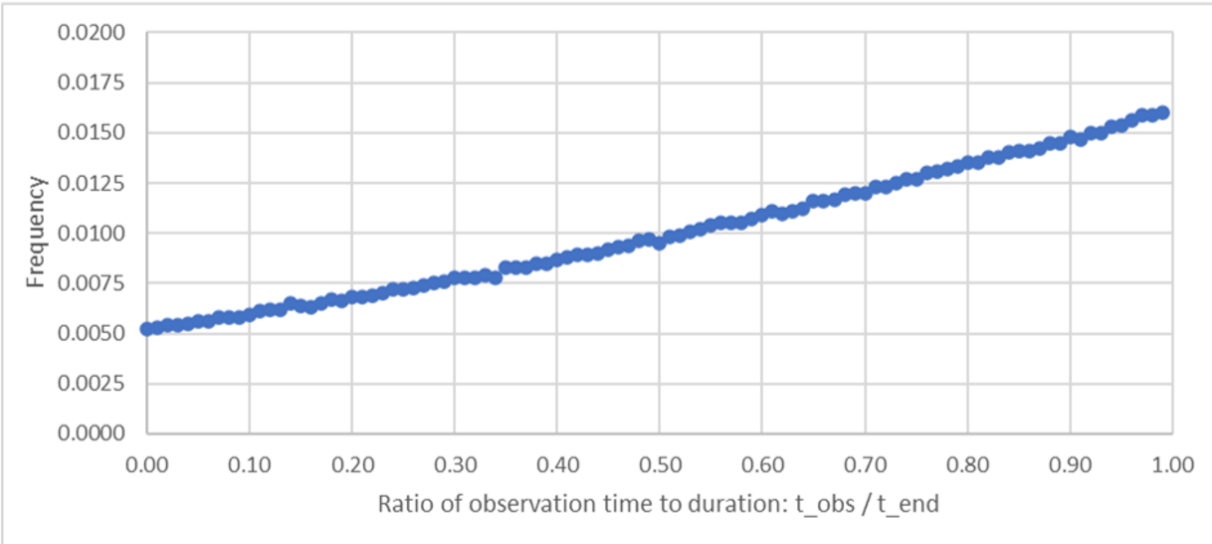

Figure 32: Histogram of the frequencies of $\frac{t_{obs}}{t_{end}}$ for 100 bins in the interval [0, 1], based on 1,000,000 Monte Carlo simulations, for Set 10 event durations distributed as U[0, 20], baby boom start times distributed as N[60, 20]. Mean of start times is 2 SDs before sample time at 100.

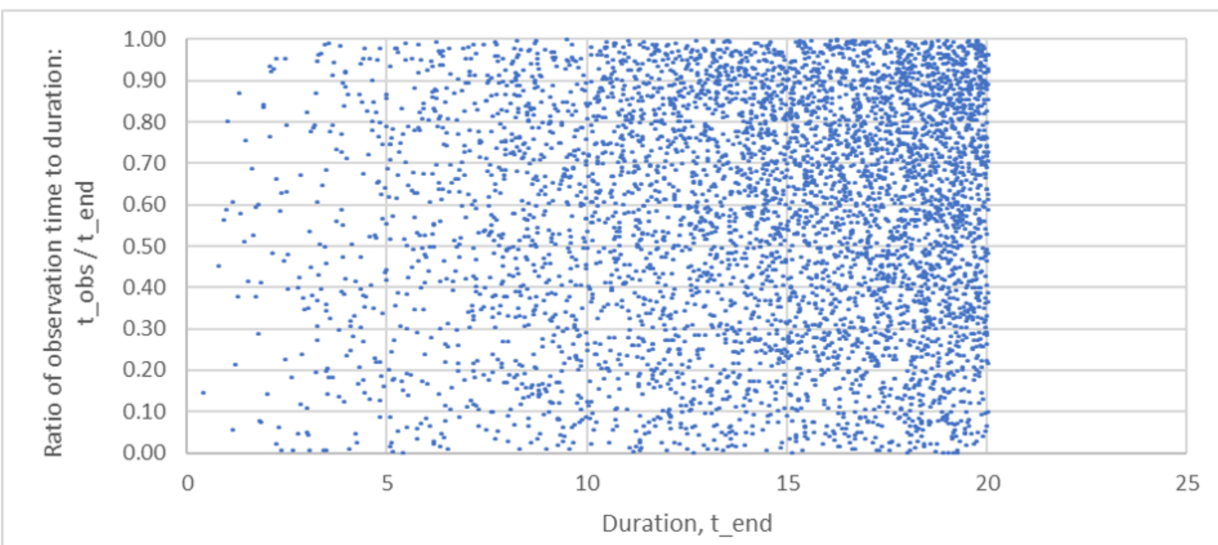

Figure 33: Scatter plot of $\frac{t_{obs}}{t_{end}}$ as a function of t_{end}, based on 5,000 Monte Carlo simulations, for Set 10 event durations distributed as U[0, 20], baby boom start times distributed as N[60, 20]. Mean of start times is 2 SDs before sample time at 100.

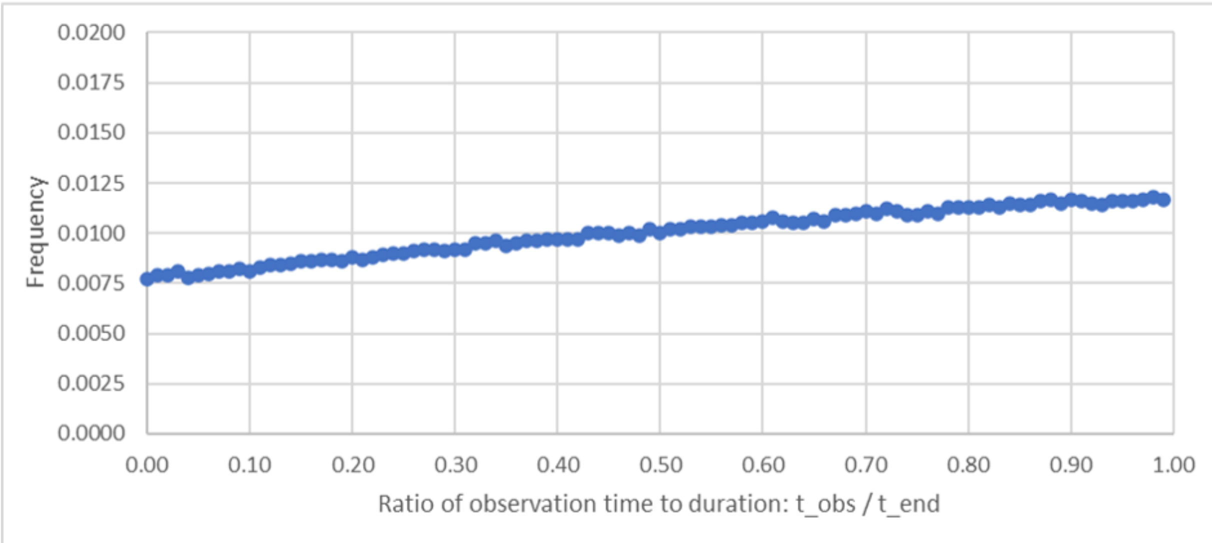

Figure 34: Histogram of the frequencies of $\frac{t_{obs}}{t_{end}}$ for 100 bins in the interval [0, 1], based on 1,000,000 Monte Carlo simulations, for Set 11 event durations distributed as U[0, 20], baby boom start times distributed as N[80, 20]. Mean of start times is 1 SD before sample time at 100.

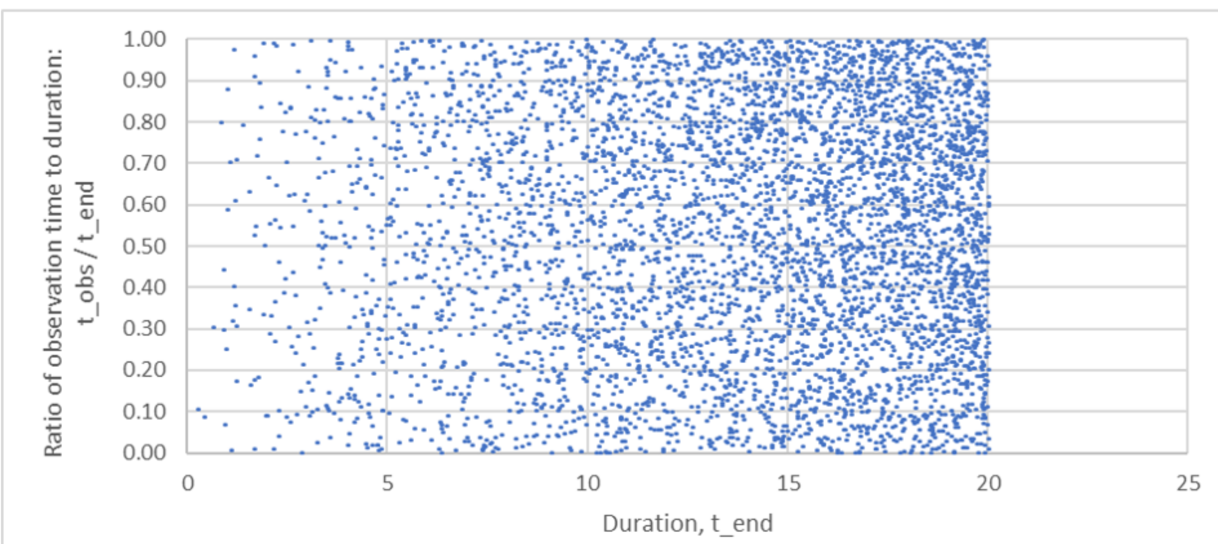

Figure 35: Scatter plot of $\frac{t_{obs}}{t_{end}}$ as a function of t_{end}, based on 5,000 Monte Carlo simulations, for Set 11 event durations distributed as U[0, 20], baby boom start times distributed as N[80, 20]. Mean of start times is 1 SD before sample time at 100.

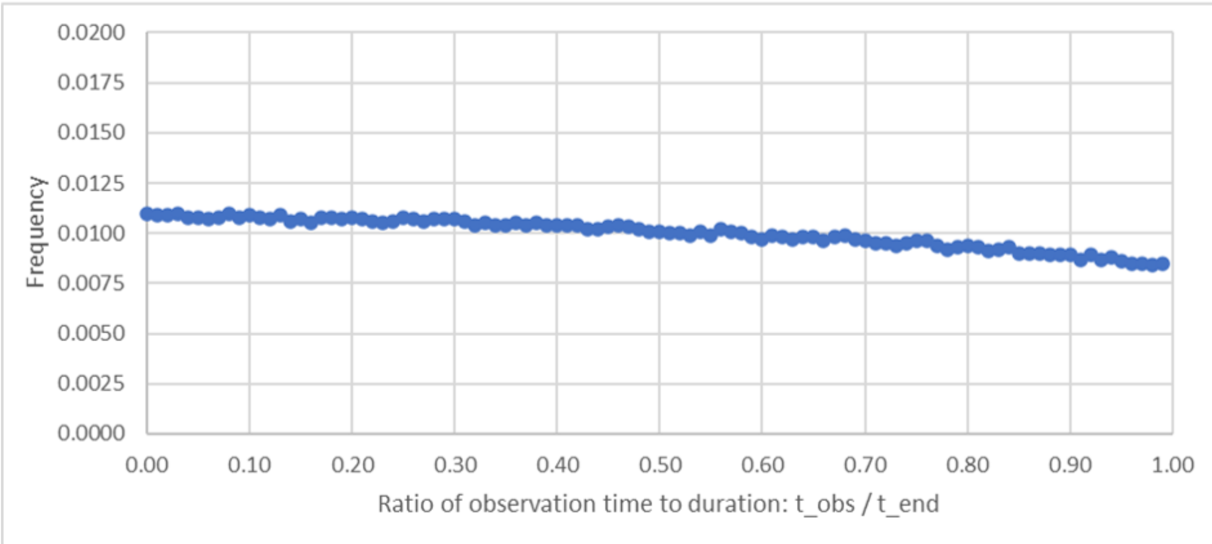

Figure 36: Histogram of the frequencies of $\frac{t_{obs}}{t_{end}}$ for 100 bins in the interval [0, 1], based on 1,000,000 Monte Carlo simulations, for Set 12 event durations distributed as U[0, 20], baby boom start times distributed as N[100, 20]. Mean of start times is at sample time at 100.

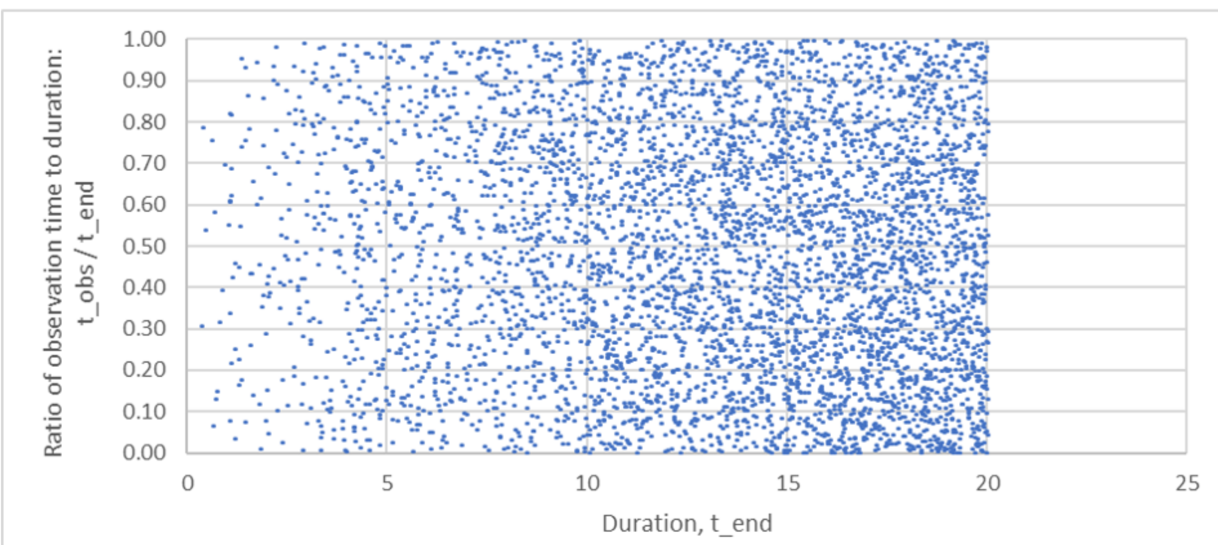

Figure 37: Scatter plot of $\frac{t_{obs}}{t_{end}}$ as a function of t_{end}, based on 5,000 Monte Carlo simulations, for Set 12 event durations distributed as U[0, 20], baby boom start times distributed as N[100, 20]. Mean of start times is at sample time at 100.

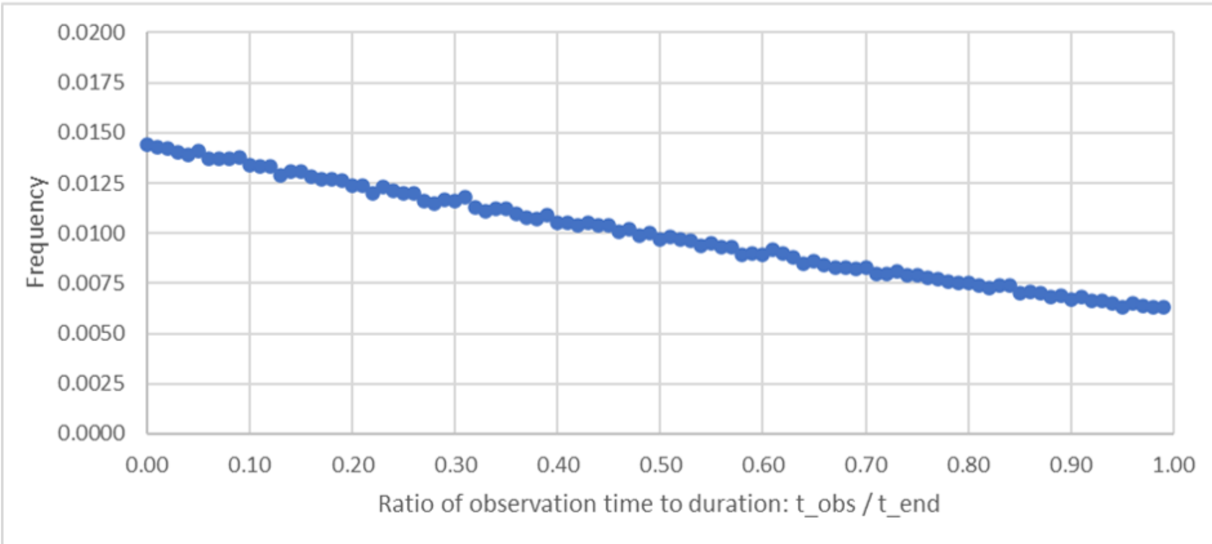

Figure 38: Histogram of the frequencies of $\frac{t_{obs}}{t_{end}}$ for 100 bins in the interval [0, 1], based on 1,000,000 Monte Carlo simulations, for Set 13 event durations distributed as U[0, 20], baby boom start times distributed as N[120, 20]. Mean of start times is 1 SD after sample time at 100.

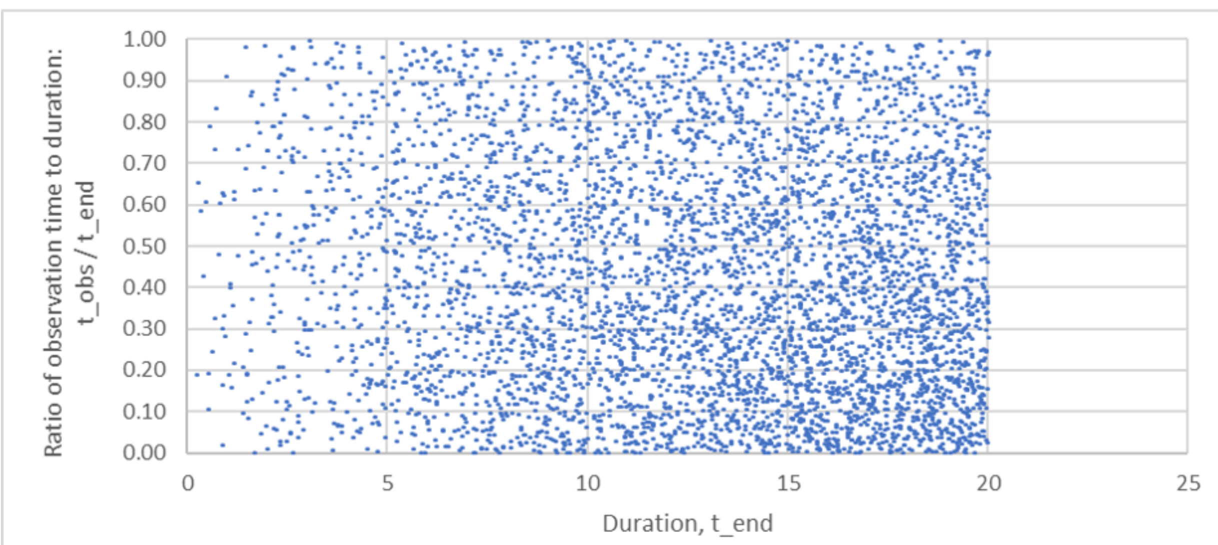

Figure 39: Scatter plot of $\frac{t_{obs}}{t_{end}}$ as a function of t_{end}, based on 5,000 Monte Carlo simulations, for Set 13 event durations distributed as U[0, 20], baby boom start times distributed as N[120, 20]. Mean of start times is 1 SD after sample time at 100.

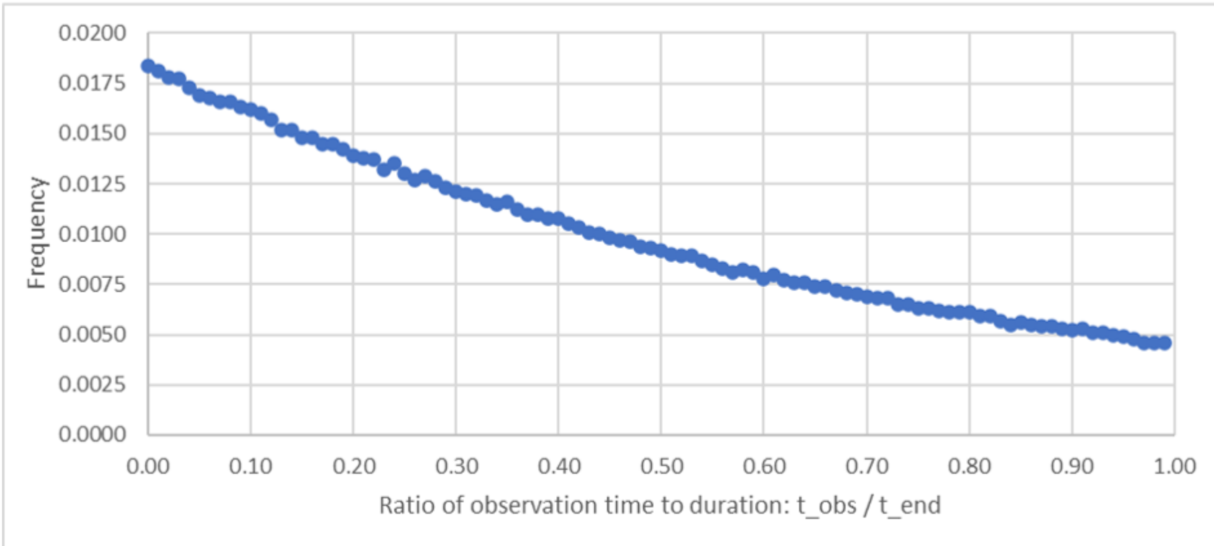

Figure 40: Histogram of the frequencies of $\frac{t_{obs}}{t_{end}}$, for Set 14 event durations distributed as U[0, 20], baby boom start times distributed as N[140, 20]. Mean of start times is 2 SDs after sample time at 100.

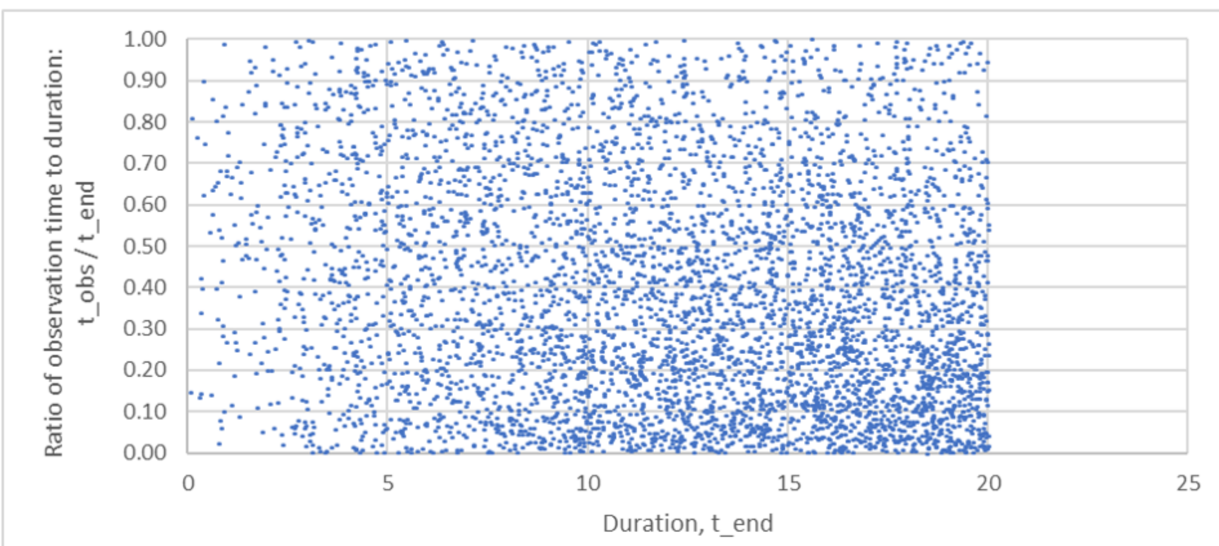

Figure 41: Scatter plot of $\frac{t_{obs}}{t_{end}}$ as a function of t_{end}, based on 5,000 Monte Carlo simulations, for Set 14 event durations distributed as U[0, 20], baby boom start times distributed as N[140, 20]. Mean of start times is 2 SDs after sample time at 100.

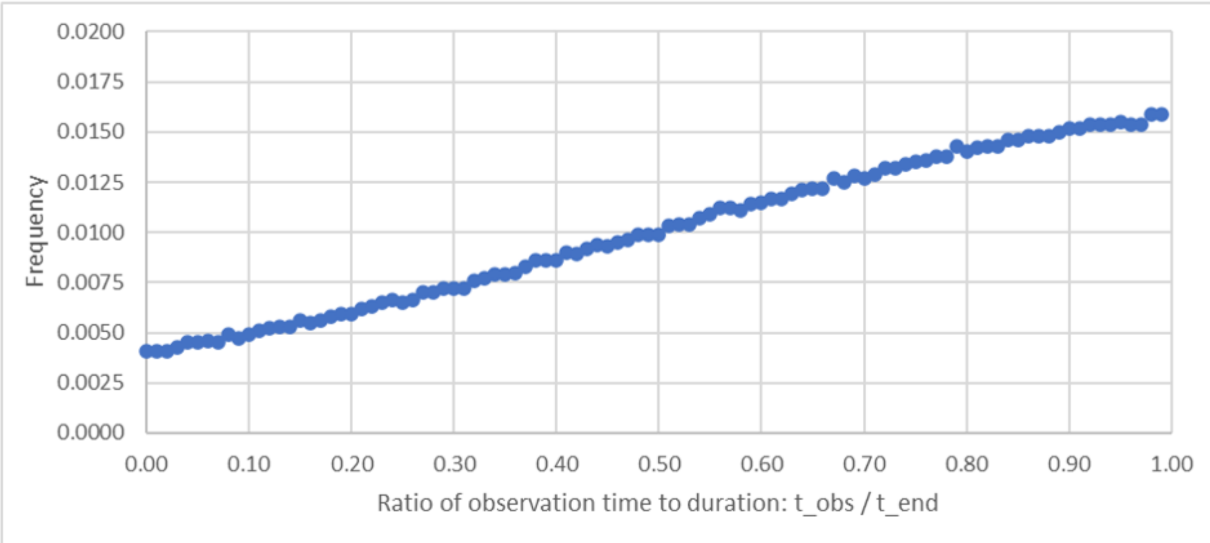

Figure 42: Histogram of the frequencies of $\frac{t_{obs}}{t_{end}}$ for 100 bins in the interval [0, 1], based on 1,000,000 Monte Carlo simulations, for Set 15 event durations exponentially distributed with MTTF = $1/\lambda = 10$, baby boom start times distributed as N[60, 20]. Mean of start times is 2 SDs before sample time at 100.

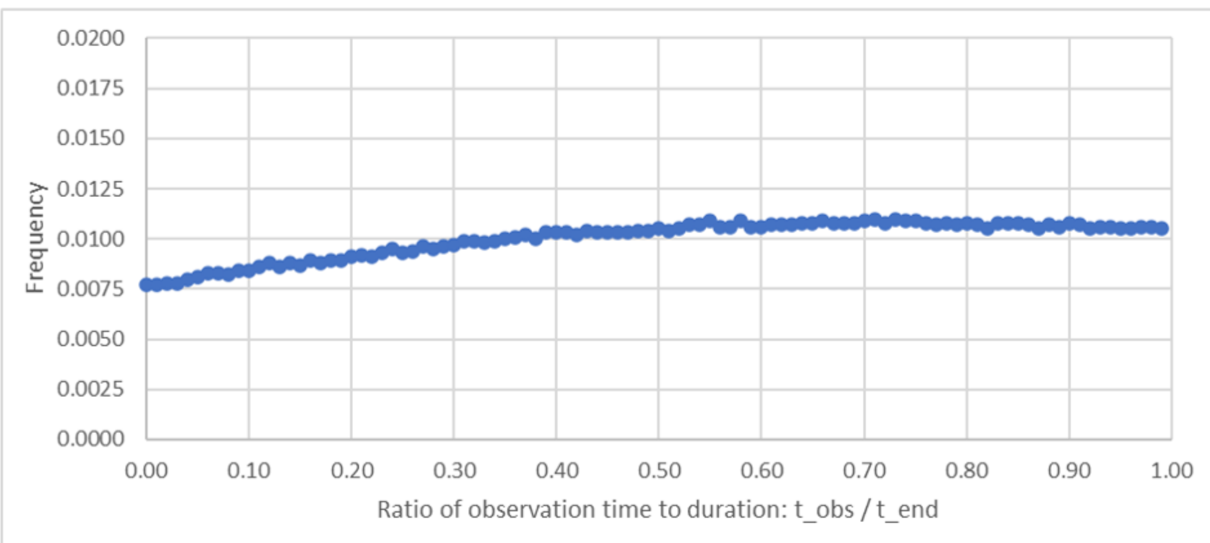

Figure 43: Histogram of the frequencies of $\frac{t_{obs}}{t_{end}}$ for 100 bins in the interval [0, 1], based on 1,000,000 Monte Carlo simulations, for Set 16 event durations exponentially distributed with MTTF = $1/\lambda = 10$, baby boom start times distributed as N[80, 20]. Mean of start times is 1 SD before sample time at 100.

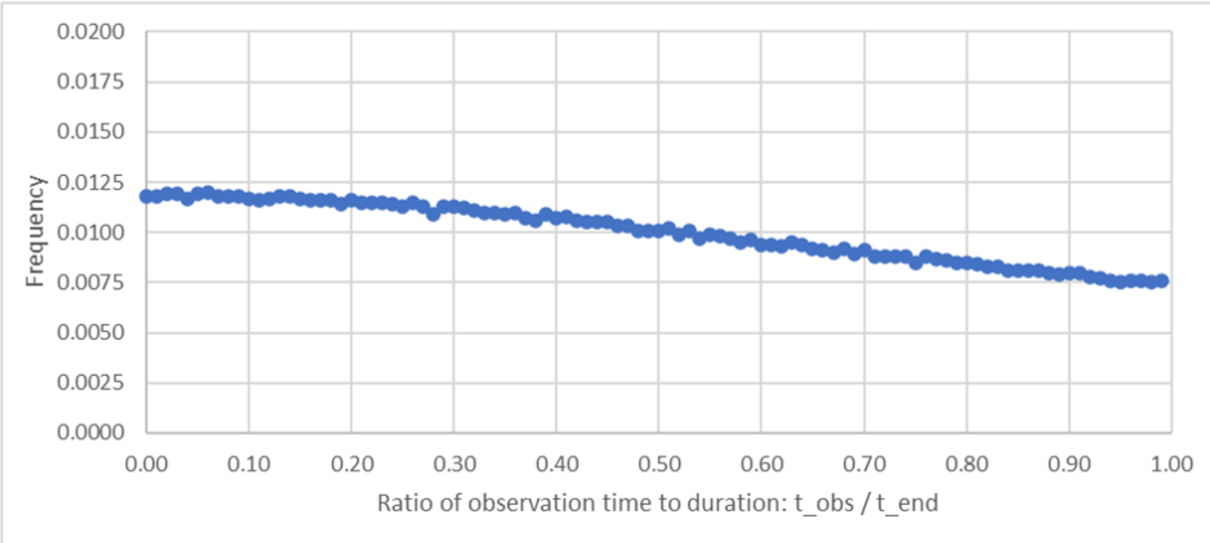

Figure 44: Histogram of the frequencies of $\frac{t_{obs}}{t_{end}}$ for 100 bins in the interval [0, 1], based on 1,000,000 Monte Carlo simulations, for Set 17 event durations exponentially distributed with MTTF = $1/\lambda$ = 10, baby boom start times distributed as N[100, 20]. Mean of start times is at sample time at 100.

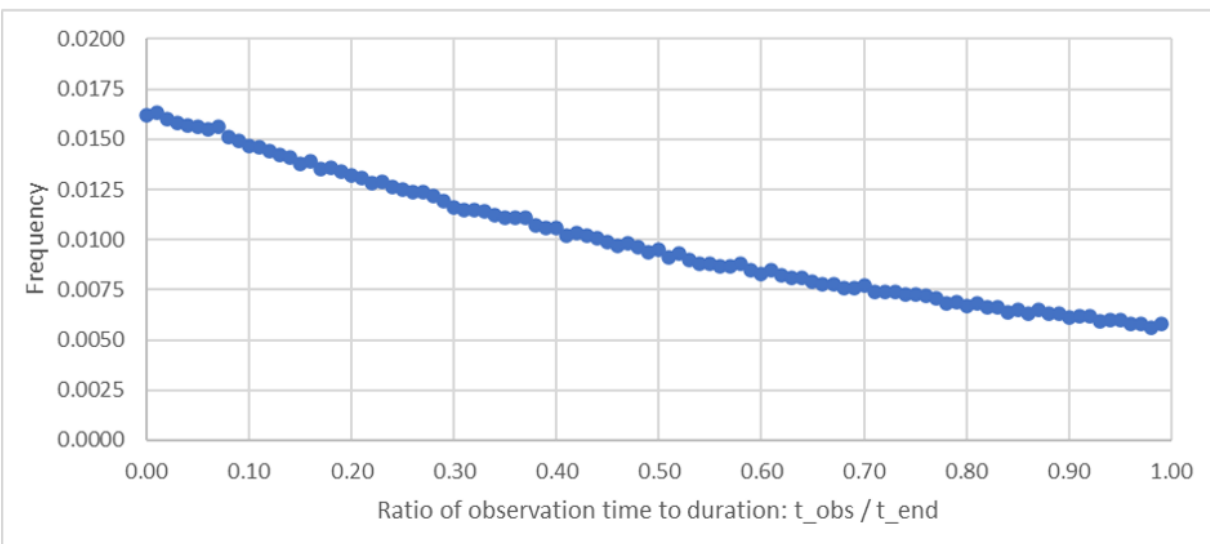

Figure 45: Histogram of the frequencies of $\frac{t_{obs}}{t_{end}}$ for 100 bins in the interval [0, 1], based on 1,000,000 Monte Carlo simulations, for Set 18 event durations exponentially distributed with MTTF = $1/\lambda$ = 10, baby boom start times distributed as N[120, 20]. Mean of start times is 1 SD after sample time at 100.

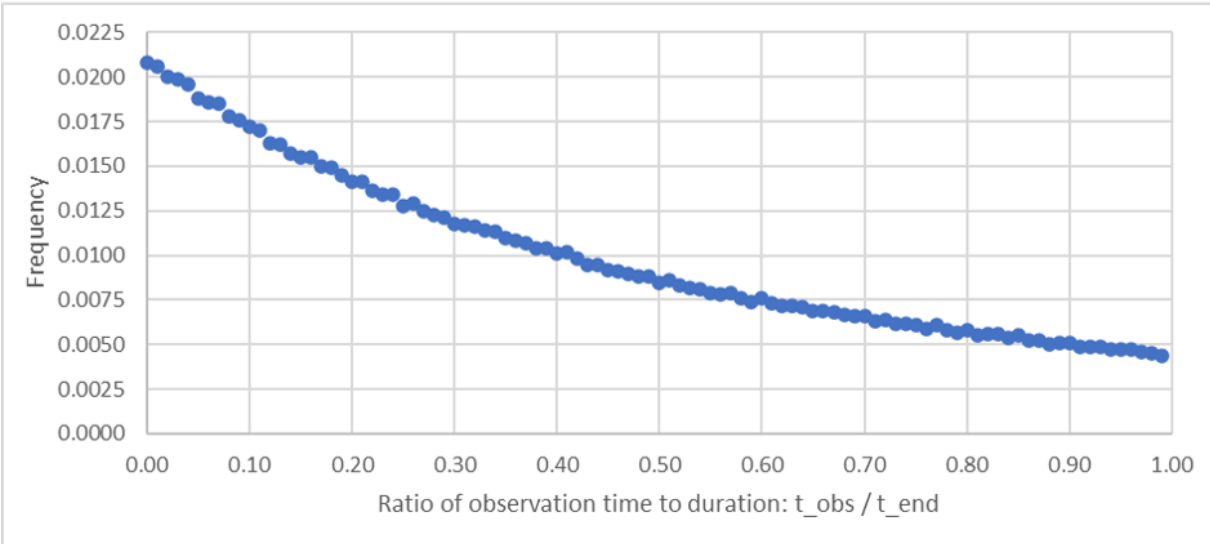

Figure 46: Histogram of the frequencies of $\frac{t_{obs}}{t_{end}}$ for 100 bins in the interval [0, 1], based on 1,000,000 Monte Carlo simulations, for Set 19 event durations exponentially distributed with MTTF = $1/\lambda = 10$, baby boom start times distributed as N[140, 20]. Mean of start times is 2 SDs after sample time at 100.

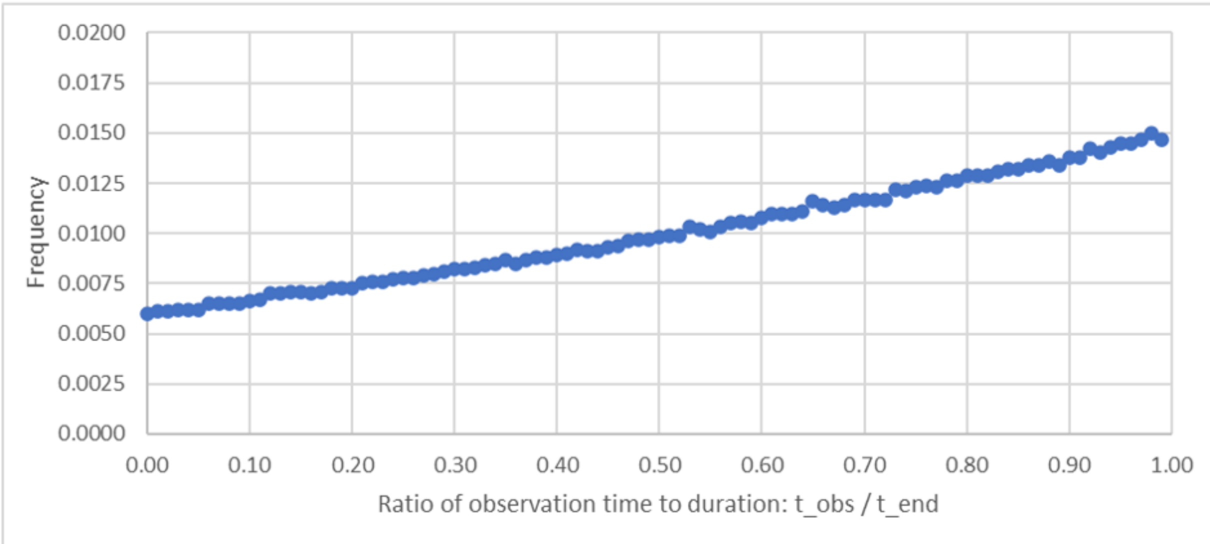

Figure 47: Histogram of the frequencies of $\frac{t_{obs}}{t_{end}}$ for 100 bins in the interval [0, 1], based on 1,000,000 Monte Carlo simulations, for Set 20 event durations distributed as N[10, 2], baby boom start times distributed as N[60, 20]. Mean of start times is 2 SDs before sample time at 100.

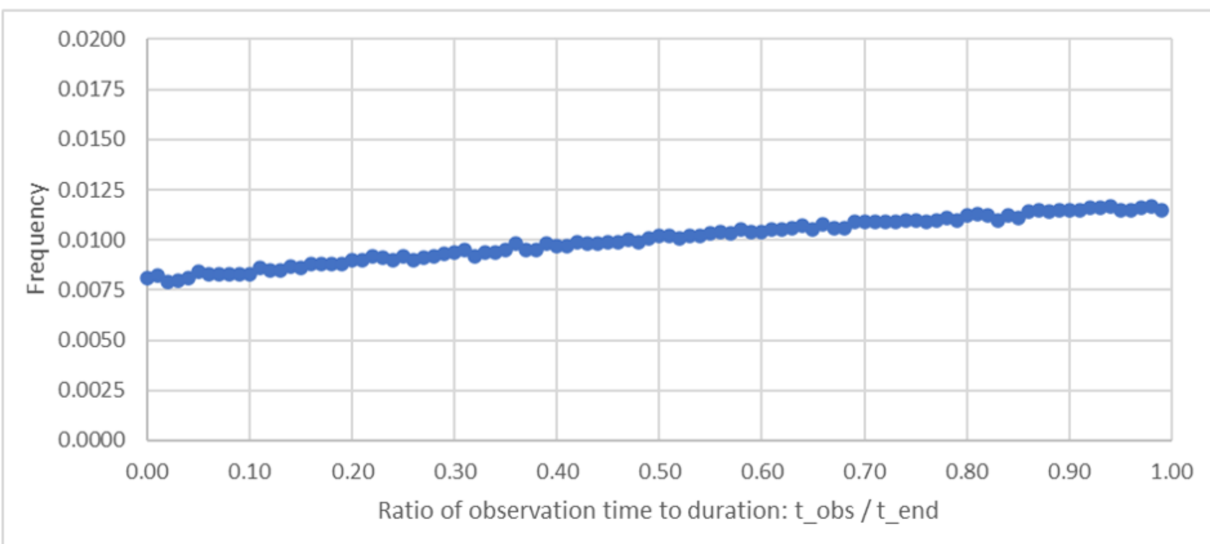

Figure 48: Histogram of the frequencies of $\frac{t_{obs}}{t_{end}}$ for 100 bins in the interval [0, 1], based on 1,000,000 Monte Carlo simulations, for Set 21 event durations distributed as N[10, 2], baby boom start times distributed as N[80, 20]. Mean of start times is 1 SD before sample time at 100.

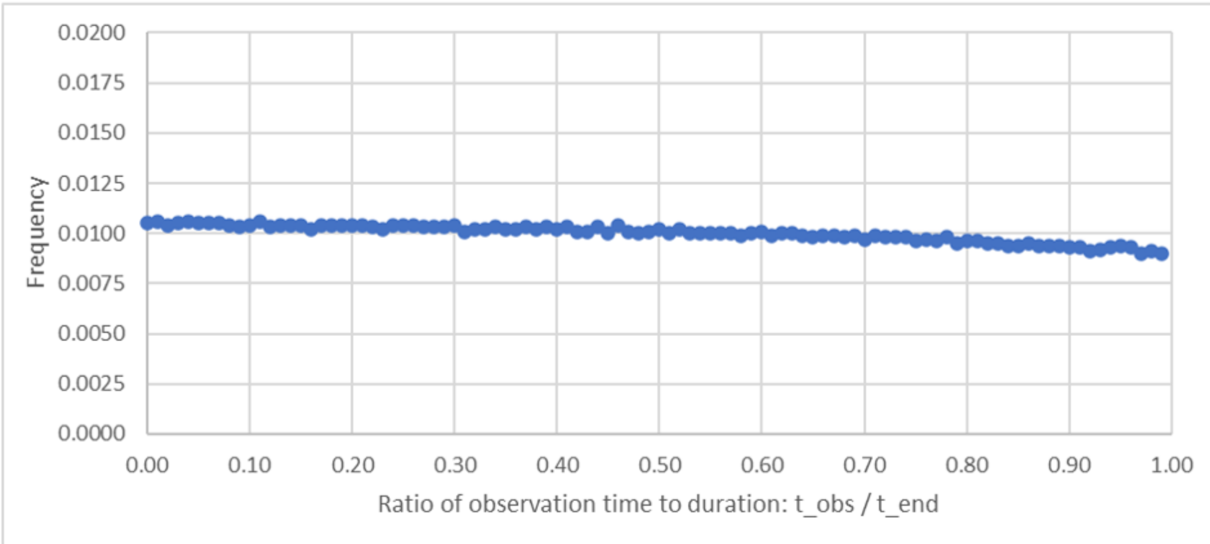

Figure 49: Histogram of the frequencies of $\frac{t_{obs}}{t_{end}}$ for 100 bins in the interval [0, 1], based on 1,000,000 Monte Carlo simulations, for Set 22 event durations distributed as N[10, 2], baby boom start times distributed as N[100, 20]. Mean of start times is at sample time at 100.

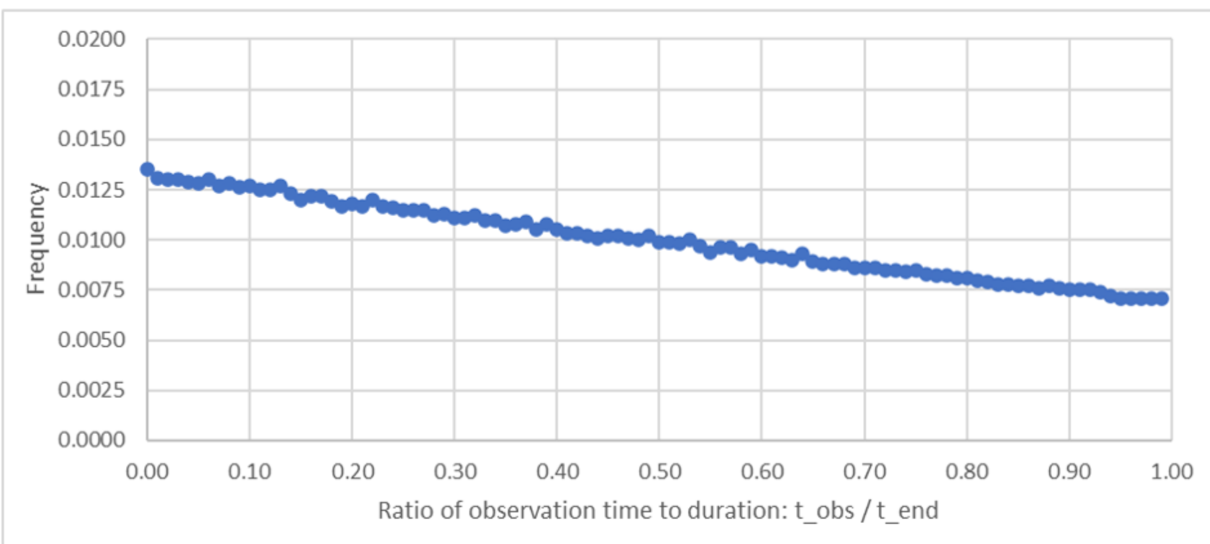

Figure 50: Histogram of the frequencies of $\frac{t_{obs}}{t_{end}}$ for 100 bins in the interval [0, 1], based on 1,000,000 Monte Carlo simulations, for Set 23 event durations distributed as N[10, 2], baby boom start times distributed as N[120, 20]. Mean of start times is 1 SD after sample time at 100.

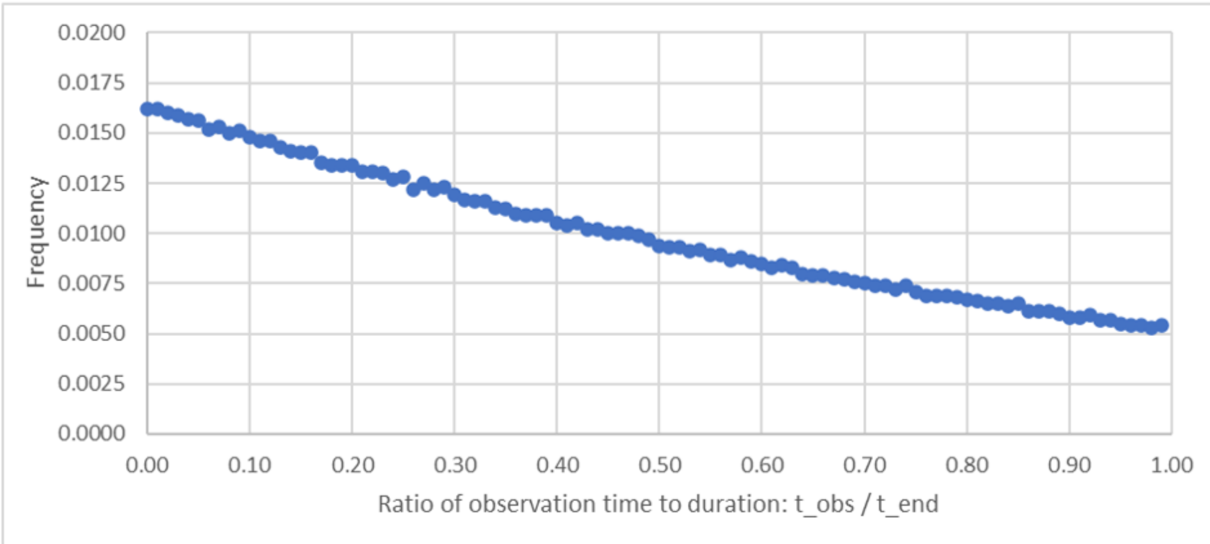

Figure 51: Histogram of the frequencies of $\frac{t_{obs}}{t_{end}}$ for 100 bins in the interval [0, 1], based on 1,000,000 Monte Carlo simulations, for Set 24 event durations distributed as N[10, 2], baby boom start times distributed as N[140, 20]. Mean of start times is 2 SDs after sample time at 100.

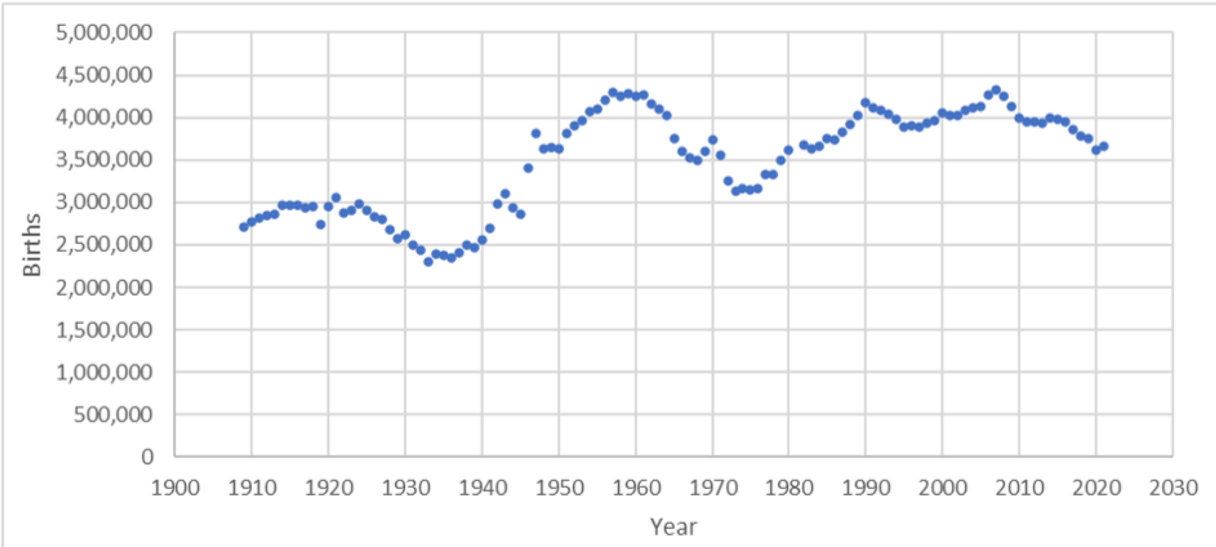

Figure 52: Birth rate data from CDC et al. with historical baby booms, used for the Set 25 Monte Carlo simulation start times to test SSA data for human lifetime distribution shown in Figure 5.

8 Continuous Recalculation of Confidence Intervals

This section continues the investigation of the consequences of violating the Copernican Principle by exploring an extreme way of doing so.

Suppose that, instead of making a single observation at no special time in the duration of the event, the observer were to *continuously observe*[34] an ongoing event from its start time to its end time, and perform a *continuous recalculation* of the LB and UB of a CI using one of the available methods. At the inevitable end of the event duration, the quality of these CIs can be assessed *retrospectively*: one can look back and see how often the sequence of CIs bounded the eventual duration.

In practice, "continuous recalculation" would probably mean "relatively frequent recalculation" on a regular or random schedule for refreshes of the LB and UB of the CI as the event duration proceeds to its unknown end time.

For convenience in the following two examples of observing an ongoing event, let the true duration of this event be 1 (where the unit of time is arbitrary), although the duration is unknown to the observer during the event.

If the observer were to estimate the total duration of the event based on observing that the event is ongoing at $t_{obs} = 0.2000$, using Gott's (5), then the LB on t_{end} would be 0.2051 and the UB on t_{end} would be 8.0000. The true $t_{end} = 1$, when the event's finale eventually occurs, will be found to have been between these bounds.

On the other hand, at the much earlier observation time of $t_{obs} = 0.0200$, the observer would have obtained the LB of 0.0205 and UB of 0.8000 on the total duration. At the event's end time, it would be seen that the LB had been correct but the UB had not.

[34] Depending on one's viewpoint, it could be argued that, instead of being the most extreme way of violating the Copernican Principle, it is the most drastic way of satisfying it. That is, an argument could be made that, if all times during the event are sampled, then none are special.

This is generalized as follows. For any t_{end}, based on (5), Gott's LB is correct when

$$\begin{aligned} t_{end} \;&\geq\; (1+\rho_{LB})\, t_{obs} \\ &\text{or, for a 95\% CI:} \\ \frac{t_{obs}}{t_{end}} &\leq \frac{39}{40} = 0.975 \;\; . \end{aligned} \tag{12}$$

Similarly, Gott's UB is correct when

$$\begin{aligned} t_{end} \;&\leq\; (1+\rho_{UB})\, t_{obs} \\ &\text{or, for a 95\% CI:} \\ \frac{t_{obs}}{t_{end}} &\geq \frac{1}{40} = 0.025 \;\; . \end{aligned} \tag{13}$$

Putting bounds (12) and (13) together, Gott's 95% CI is correct when

$$0.025 \;\leq\; \frac{t_{obs}}{t_{end}} \;\leq\; 0.975 \;\; . \tag{14}$$

Expressions (12), (13), and (14) of course merely circle back to Gott's original argument embodied in (2), (3), (4), and (5) and restate them for continuous recalculation of 95% CIs for the event duration, when it is eventually known, as:

- Gott's LB will be found to have been correct for the *first* 97.5% of the event duration.
- Gott's UB will be found to have been correct for the *last* 97.5% of the event duration.
- Gott's LB and UB both will be found to have been correct for the *middle* 95% of the event duration.

This is illustrated in Figure 53 where the three step functions show where the LB, UB, and CI are correct for regions of an event duration expressed as $\frac{t_{obs}}{t_{end}}$.

Continuous recalculation of Gott's CIs given by (4) or (5) was tested for continuous populations using the C program that performed the Monte Carlo simulations described in Section 6. These tests used the first 1,000 out of the 1,000,000 Monte Carlo simulations performed for each set. For each simulation in a set, Gott's CI was calculated at 9,999 values of t_{obs} from $\frac{1}{10{,}000} \times t_{end}$ to $\frac{9{,}999}{10{,}000} \times t_{end}$ in increments of $\frac{1}{10{,}000} \times t_{end}$.[35]

For a given set, the *overall probability* of continuous recalculation of Gott's CIs being correct was estimated by the fraction of times the recalculated CI for each of the 9,999 values of t_{obs} enclosed the actual value of the duration for the 1,000 Monte Carlo simulations.

By the argument used in this section, Gott's 95% CI should be correct whenever (14) holds. As expected, it was found that the overall probability of continuous recalculation of Gott's CIs being correct was consistently 0.950 for *every* individual Monte Carlo simulation for the continuous populations for each of the Sets 1 through 9.

The conditional survival CIs were also tested using the same approach for comparison to the findings above for Gott's CIs.

In contrast to these consistent results for continuous recalculation of CIs using Gott's method, those for the conditional survival CIs showed considerable variation. Although the *average* overall probability of continuous recalculation of the conditional survival CIs being correct tended closely toward the target value of 0.950, the overall probabilities of being correct for *individual* simulations could be quite different (higher or lower) from that target value.

Some examples follow for the tests of continuous recalculation of the conditional survival CIs.

For Set 1, using event durations uniformly distributed as U[0, 20], continuous recalculation of the CIs had an average overall probability of being correct of 0.948, but a 2.5%ile value of 0.481 and a 5%ile of 0.742. The 95%ile and the 97.5%ile were 0.999. The minimum value was 0.025.

[35] This set of 9,999 values was chosen so as to avoid the endpoints of $t_{obs} = 0$ and $t_{obs} = t_{end}$.

For Set 2, using events having exponentially distributed durations with an MTTF of $1/\lambda = 10$, continuous recalculation of the CIs had an average overall probability of being correct of 0.952. The 2.5%ile value was 0.629, the 5%ile was 0.754, the 95%ile was 0.992, and the 97.5%ile was 0.993. The minimum value was 0.140.

For Set 3, using events having durations normally distributed as N[10, 5], continuous recalculation of the CIs had an average overall probability of being correct of 0.952. The 2.5%ile value was 0.457, the 5%ile was 0.674, and the 95%ile and 97.5%ile were 0.997. The minimum value was 0.

For the conditional survival method, some sets demonstrated instances of individual Monte Carlo simulations having very low overall probabilities of continuous recalculation of CIs being correct. Set 3 (as noted above) as well as Sets 4, 5, 8, and 9 were seen to have some minimum values of 0. Set 7 nearly did so with a minimum value of 0.038. Many of the baby boom sets also had minimum values of 0 or close to 0.

These results varied, but only slightly, when the sets were run repeatedly with different random seeds.

The circumstances under which the conditional survival CIs performed poorly were investigated. For Set 1, Figure 54 provides a scatter plot of the overall probability of being correct as a function of t_{end} for the first 1,000 out of the 1,000,000 Monte Carlo simulations performed. It can be seen that the smaller and larger durations resulting from that distribution had the smallest overall probabilities of being correct. This pattern was closely matched in the same kind of scatter plots for the other distributions tested.

To eliminate the possibility that this behavior at the extremes of the distributions might have been an artifact of estimating the reliability function on the basis of the same number of samples as the number of Monte Carlo simulations, the experiments were re-run using the much larger number of 10,000,000 samples of the distribution, again independently of the Monte Carlo simulations, to estimate $R(t)$. The numerical differences from the results reported above were slight, and the graphs were almost indistinguishable. Thus, one can reasonably conclude that this behavior of continuous recalculation of the conditional survival CIs is real.

Summarizing these findings, Gott's CIs yielded overall probabilities of being correct of 0.950 when continuously recalculated over the total duration for each of the simulated events for each of the sets tested. In contrast, while the conditional survival method achieved an average overall probability of being correct of 0.950, in many cases the CIs for individual events were only seldom correct or never correct over the total duration.

Furthermore, it has been established that the LB of Gott's 95% CI is correct for the first 97.5% of an unknown event duration and that the LB on the duration of an event representing a critical function can be considered a measure of trustworthiness. For these reasons, continuous recalculation of Gott's LB over a critical function's lifetime could be interpreted as a running estimate of the current statistical level of trust that the function will continue to operate for an additional specific amount of time.

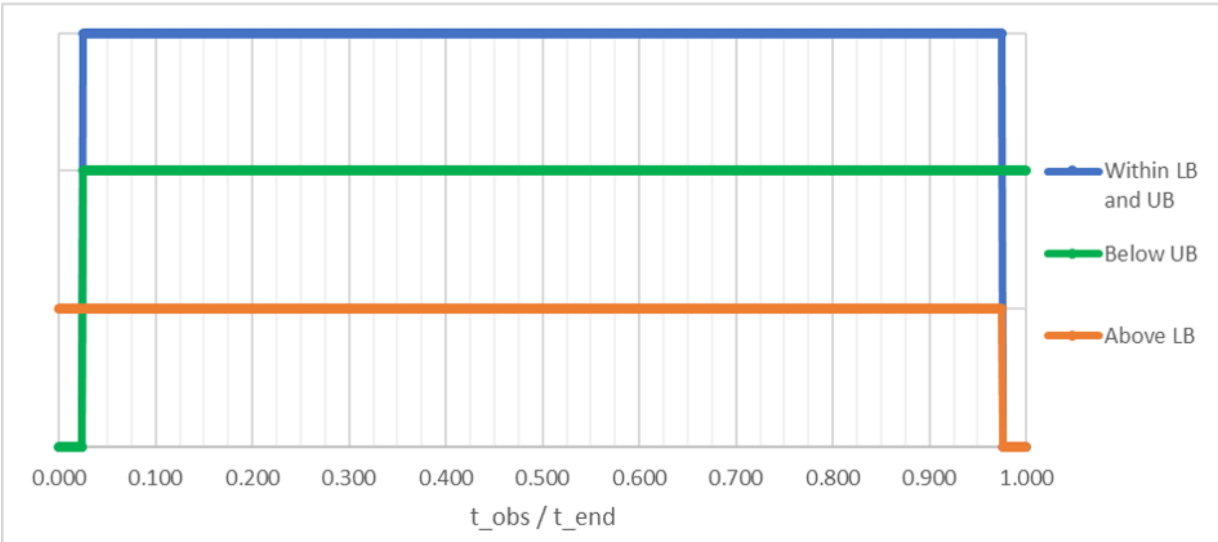

Figure 53: Continuous sampling of an event, from its start to its end, while continuously recalculating Gott's LB and UB. The steps are indicator functions showing the points in the duration at which the total duration will be found to have been *within* the calculated CI, *at or above* the calculated LB, and *at or below* the calculated UB.

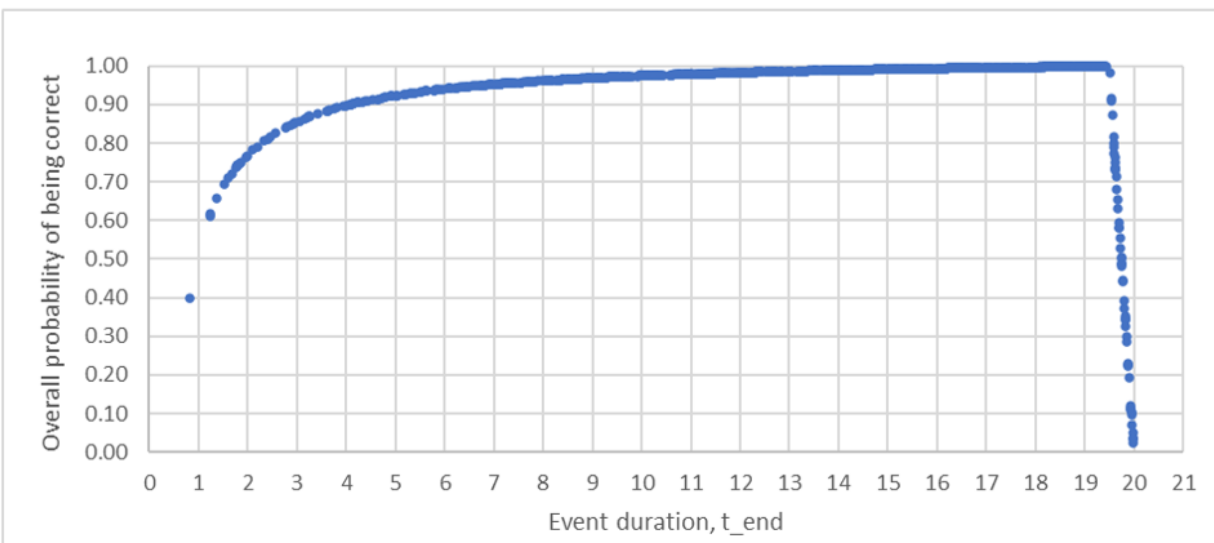

Figure 54: Scatter plot of the overall probability that conditional survival CIs are correct as a function of t_{end}, for 1,000 Monte Carlo simulations for Set 1 event durations distributed as U[0, 20] for a constant population.

9 Conclusions

This paper has tested a method proposed by J. R. Gott for obtaining confidence intervals or CIs for the remaining duration of an observed event when its past duration at that time is known to the observer. The only necessary assumption of Gott's method is that the Copernican Principle holds, here meaning that observation of an event occurs at no special time in the duration of the event.

Gott's method presumes that the reliability function $R(t)$, or other additional information that governs the event duration, or external knowledge about the end of the event is unknown to the observer. Otherwise, tighter bounds on a CI can and should be obtained using an alternative applicable method.

Most of the results presented in this paper are obtained for a model referred to as "Copernican Sample." In this model, populations of events (which might consist of only a single event) have uniformly distributed start times in some finite timespan. An observer captures an event, if it is ongoing at a given sample time, and notes its past duration. Gott's method projects that the remaining duration is between $1/39$ and 39 times the past duration at the time of the observation with 95% confidence.

Much better bounds were obtained using the known reliability functions for event durations, but Gott's method was demonstrated to work *regardless* of the underlying distribution. The advantage of Gott's method is due to the fact that, in practice, the underlying distribution may be effectively unknown or even unknowable.

While it is presumed that Gott's method is applied only when the underlying distribution is not known to the observer, it may at least be known that the only crucial portion of the duration is the time spent in active use rather than storage or standby conditions. In such cases, Gott's method, as well as any other means of calculating CIs, would be more accurate when duration is measured in terms of active use instead of strict calendar time. When operational time is accumulated such that it is proportional to calendar time, however, this distinction may be unnecessary as Gott's CIs would scale appropriately.

This paper demonstrated the application of Gott's method for several distributions and mixtures of distributions using Monte Carlo simulations consisting of 1,000,000 observed events. For constant populations, meaning steady state or uniformly distributed start times for the events, Gott's 95% CIs for the Copernican Sample model achieved that target confidence level. The tail probabilities were 2.5%, so the CIs were central as well.

This paper also explored the performance of Gott's method for an alternative although deprecated model of the Copernican Principle, referred to as "Copernican Observation." In this model, the observation time is explicitly selected uniformly within the duration of an ongoing event. While Gott's method is shown to work for this model as well, it was included only for completeness as its application in the real world outside of simulation would seem to be unrealizable.

Gott's method was also tested under conditions that violated the Copernican Principle to some extent. Populations of events were created with non-uniformly distributed "baby booms" of start times. This model generated event durations with the same underlying distributions used for the constant population experiments, but with normally distributed start times centered at the sample time as well as 1 and 2 standard deviations before and after the sample time. In spite of these divergences from the only necessary assumption of Gott's method, it was found that the confidence levels achieved by Gott's CIs were still close to the target of 95%; in these cases, however, the CIs were slightly non-central as the tail probabilities were unequal.

Another test flouted the Copernican Principle by using event durations that modeled human lifetimes with start times generated according to historical birth data; these data included human population baby booms. The results of the Monte Carlo simulations showed that Gott's CI was tolerant to this violation of the Copernican Principle as well, and achieved a confidence level close to 95% although the CI was again slightly non-central.

The conditional survival method, which uses a known reliability function for the event durations, was the baseline for comparison to Gott's CIs. The reliability function $R(t)$ was obtained by storing 1,000,000 independent samples of the underlying distribution of the event durations in a lookup table. The conditional survival method also yielded 95% CIs and, as expected, their bounds were much tighter due to the additional information.

It was noted, at least in these experiments, that the lower bounds provided by Gott's two-sided 95% CIs tended to be very close to those of the conditional survival CIs. As the lower bound on the remaining duration that a critical function will operate can be a useful measure of trustworthiness, Gott's lower bound may provide such an assurance even in the absence of information about that function's underlying distribution for time to failure.

Continuous recalculation of the CIs was tested over the entirety of the durations of individual events, which can be viewed as an extreme violation of the Copernican Principle. At the end of any event, an observer would retrospectively find that continuous recalculation of Gott's CIs had an overall probability of being correct that was equal to the target confidence level of 95%. "Continuous recalculation" in practice would probably mean "relatively frequent recalculation."

For the conditional survival method, on the other hand, it was found that the overall probability of continuous recalculation of the CIs being correct averaged approximately 95% over all event durations, but could be much lower for the smaller and larger event durations resulting from that distribution.

It is argued that continuous recalculation of Gott's LB over a critical function's lifetime may be interpreted as a running estimate of the current statistical level of trust that a currently operating function will continue to do so for an additional specific amount of time.

This paper makes no recommendations as to the applicability of Gott's method for calculating CIs for the remaining duration or total duration of an event for any specific situation. As pointed out in the Introduction, variation is the norm, and so it is important to remember that no CI is a guarantee; for example, any 95% CI is expected to have about 1 in 20 of its outcomes outside that interval. The purpose of this paper is to provide researchers with empirically based data upon which to further determine the fitness for use of the Copernican Principle for estimating these bounds, whether in the principle's strict sense of observation taking place at no special time during the event, or in the relaxed cases studied here.

Acknowledgment

Dr. E. Paul Ratazzi (AFRL) provided invaluable technical and procedural advice during the performance of this work.

References

[Bondi (1948)] H. Bondi and T. Gold, "The Steady-State Theory of the Expanding Universe," *Monthly Notices of the Royal Astronomical Society*, Vol 108(3), 1948.

[Brown (2001)] L. Brown, T. Cai, and A. DasGupta, "Interval Estimation for a Binomial Proportion," *Statistical Science*, Vol 16(2), 2001.

[Caves (2008)] C. M. Caves, "Predicting Future Duration from Present Age: Revisiting a Critical Assessment of Gott's Rule," *arXiv preprint arXiv:0806.3538*, 2008.

[CDC (2023)] Centers for Disease Control (CDC), Infoplease, and Statistica, "Live Birth Data," https://www.cdc.gov/nchs/data/statab/t1x0197.pdf, https://www.infoplease.com/us/population/live-births-and-birth-rates-year, https://www.statista.com/statistics/195908/number-of-births-in-the-united-states-since-1990/, 2023.

[Cowles (1982)] M. Cowles and C. Davis, "On the Origins of the .05 Level of Statistical Significance," *American Psychologist*, Vol 37(5), 1982.

[Debany (2023)] W. H. Debany, "Improved Methods for Exact Confidence Intervals When Sampling With and Without Replacement for Rare Events," DTIC #AD1208335, https://apps.dtic.mil/sti/citations/trecms/AD1208335, Aug 2023.

[Downey (2019)] A. Downey, "The Inspection Paradox is Everywhere," *Towards Data Science*, https://towardsdatascience.com/the-inspection-paradox-is-everywhere-2ef1c2e9d709, Aug 2019.

[Edney (2004)] J. Edney and W. A. Arbaugh, *Real 802.11 Security: Wi-Fi Protected Access and 802.11i*, Pearson Education, 2004.

[Gale (1981)] G. Gale, "The Anthropic Principle," *Scientific American*, Vol 245(6), 1981.

[Gott (1993)] J. R. Gott, "Implications of the Copernican Principle for Our Future Prospects," *Nature*, Vol 363, 1993.

[Hassani (2016)] M. Hassani, R. Kippen, and T. Mills, "Probability: A Matter of Life and Death," *Australian Senior Mathematics Journal*, Vol 30(1), 2016.

[Heyman (1982)] D. P. Heyman and M. J. Sobel, *Stochastic Models in Operations Research Vol I: Stochastic Processes and Operating Characteristics*, McGraw-Hill, 1982.

[Hieke (2015)] S. Hieke, M. Kleber, C. König, M. Engelhardt, and M. Schumacher, "Conditional survival: A Useful Concept to Provide Information on How Prognosis Evolves Over Time," *Clinical Cancer Research*, Vol 21(7), 2015.

[Larson (1981)] R. C. Larson and A. R. Odoni, *Urban Operations Research*, Prentice-Hall, 1981, Sec 7.1.3 "Generating Samples from Probability Distributions," available online from Massachusetts Institute of Technology at https://web.mit.edu/urban_or_book/www/book/chapter7/7.1.3.html.

[Merriam-Webster (2005)] *Merriam-Webster's Collegiate Dictionary*, 11th Ed, Merriam-Webster, 2005.

[MIL-HDBK-217F (1990)] *Military Handbook: Reliability Prediction of Electronic Equipment (MIL-HDBK-217F)*, https://www.quanterion.com/wp-content/uploads/2014/09/MIL-HDBK-217F.pdf, Jan 1990.

[Monton (2006)] B. Monton and A. Kierland, "How to Predict Future Duration from Present Age," *The Philosophical Quarterly*, Vol 56(222), 2006.

[Nature (1994)] Comments by S. N. Goodman, A. L. Mackay, and P. Buch on J. R. Gott's "Implications of the Copernican Principle for Our Future Prospects" [Gott (1993)], *Nature*, Vol 368, Mar 1994.

[Pires (2008)] A. M. Pires and C. Amado, "Interval Estimators for a Binomial Proportion: Comparison of Twenty Methods," *REVSTAT-Statistical Journal*, Vol 6(2), 2008.

[Pham (2007)] H. Pham, *System Software Reliability*, Springer Science & Business Media, 2007.

[Pisaturo (2009)] R. Pisaturo, "Past Longevity as Evidence for the Future," *Philosophy of Science*, Vol 76(1), 2009.

[Press (2002)] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C++: The Art of Scientific Computing*, Cambridge University Press, 2002.

[Rohatgi (1979)] V. K. Rohatgi, *An Introduction to Probability Theory and Mathematical Statistics*, John Wiley & Sons, New York, 1979.

[Rudnicki (1989)] K. Rudnicki, "The Importance of Cosmological Principles for Research in Cosmology," *Apeiron*, Vol 4(1), 1989.

[Siewiorek (1982)] D. P. Siewiorek and R. S. Swarz, *The Theory and Practice of Reliable System Design*, Digital Press, 1982.

[Smith (2016)] D. J. Smith and K. G. L. Simpson, *The Safety Critical Systems Handbook: A Straightforward Guide To Functional Safety: IEC 61508 (2010 Edition), IEC 61511 (2016 Edition) & Related Guidance*, 4^{th} Ed, Elsevier, 2016.

[SSA (2023)] Social Security Administration, "Actuarial Life Table: 2020 period life table for the Social Security area population," https://www.ssa.gov/oact/STATS/table4c6.html, 2023.

[Varshney (2022)] K. R. Varshney, *Trustworthy Machine Learning*, Independently Published, 2022.

[Viadinugroho (2021)] R. A. A. Viadinugroho, "Generate Random Variable Using Inverse Transform Method in Python," *Towards Data Science*, https://towardsdatascience.com/generate-random-variable-using-inverse-transform-method-in-python-8e5392f170a3, Mar 2021.

[Weisstein (2019)] E. W. Weisstein, "Box-Muller Transformation," *MathWorld—A Wolfram Web Resource*: https://mathworld.wolfram.com/Box-MullerTransformation.html, original date of article unknown.