Properties of Rare Fail-States and Extreme Values of TTF in a Markov Chain Model for Software Reliability

Michael G. Thomason and James A. Whittaker

April 19, 1999

Abstract

Numerous probability distributions are used in the estimation of software reliability. Typically, these distributions arise from assumptions together with actual success/failure data collected during testing or field-usage. Statistics of the extremes has been applied by Kaufman et al for software reliability analysis when failure is an infrequent, unlikely occurrence—a so-called rare event. When applicable, these statistics give powerful results concerning limiting distributions without the assumption of initial distributions often imposed in existing reliability models. This paper combines (i) results in rare events and extreme values with (ii) a finite-state, discrete-parameter, recurrent Markov chain which incorporates both the failures as rare events (as transitions to a rare fail-state) and usage of the software between failures (as transitions among ordinary usage-states not involving the fail-state). Four distributions arise naturally as approximations for the chain, namely: the “Poisson law of small numbers” gives an explicit error-bound on a Poisson Approximation for count of occurrences of a rare fail-state in long intervals of software usage; the interoccurrence time of the rare fail-state (the time-to-failure or TTF) has an approximately exponential distribution; and the Weibull and Gumble distributions, respectively, are the limiting distributions of the minimum and maximum values (the extreme values) in independent samples of the TTF. Results are illustrated by examples, including $\chi^2$ goodness-of-fit tests for samples and approximate distributions.

Keywords: exponential, extreme values, Gumbel, Markov chain, Poisson, rare event, software reliability, Weibull

1 Introduction

The field of software reliability deals with the probabilistic estimation and prediction of software quality by combining stochastic models with statistical analysis of testing and field-
usage data. Such data typically includes \textit{successful executions} in which the software’s functionality was exercised without deviation from the specification, and \textit{unsuccessful or faulty runs} in which the software performed one or more of its tasks incorrectly. Performing a task incorrectly is called a \textit{failure event}. A software reliability model encapsulates observed success/failure data or other estimates of component reliability within the framework of probability models in order to predict patterns of future performance. The probability distributions for the number of failures within a specified time span, the failure interoccurrence time, and other relevant random variables are often adapted from hardware reliability theory [9] or justified empirically [1, 17].

\textit{Rare events and extreme values} are topics in probability and statistics with application in several fields of science and engineering (cf. [3, 5, 6, 13, 20]). When the necessary conditions are met, these topics lead to well-defined distributions as approximate probability laws that are relevant in restricted, but important, areas of software reliability analysis. \textit{Statistics of the extremes} has been applied to software reliability by Kaufman \textit{et al} [14, 15] for an analysis of rare event data (for infrequent, unlikely failures) without requiring \textit{a priori} knowledge of its distribution. An application for safety critical software systems is described in [14].

By calling failure a \textit{rare event} [3, 13], we mean here that its probability of occurrence is greater than zero but smaller by at least several orders of magnitude than non-failure events in software execution intervals. In this situation, the mean time-to-failure (MTTF) is a large number. \textit{Extreme values} [5, 6, 13] are the minimum and maximum values among independent\footnote{Independent means \textit{stochastically independent} throughout this paper.}, identically distributed (\textit{iid}) random variables; thus, analysis of the time-to-failure (TTF) may look not only at the MTTF but also at extreme variation above and below this mean value in multiple, \textit{iid} executions of software. This paper concentrates on describing the TTF when failure is a rare event.

Intuitive definitions and convenient computations are attributes of Markov chains for work with rare events and extreme values. In this paper, a discrete-parameter, finite-state Markov chain [24] is used to represent both software failures (as transitions to a rare fail-state) and usage of the software between failures (as transitions among the ordinary usage states not involving the fail-state). Other work with Markov chains in software reliability includes computation of the sensitivity of system reliability to a module’s reliability via a discrete-parameter Markov chain [8], generation of test cases for software systems based on a birth-death Markov chain [2], and versions of discrete- and continuous-parameter Markov chains as structural models in software reliability prediction [12]. Markov chains can be developed hierarchically for different levels of representation of software [23], can be used to model the interconnection of components [12], facilitate bookkeeping with frequency counts in empirical data [24], and offer an intuitive definition of rare events in terms of a recurrent chain’s stationary probability distribution. Generic results in rare events and extreme values are not restricted to Markov chains and may, with suitable formulation, apply to other stochastic models [3, 13]. Other models are not explored here.

The requirement that failure be a rare event is not met in all software development activity; however, it is often the case that software evolves into a phase in which it rarely fails but its past history of failure with nonzero relative frequency, or an estimate of less-than-perfect reliability of some component(s), should not be completely discounted. For
example, failures may be rare events in situations such as the following:

- Post-release “beta” versions of software, from which failures have been removed previously, can lead to failure as a rare event. If users of beta versions build up significant histories of successful execution intervals and few faulty runs, then failures may be rare events when compared with many non-failure execution intervals.

- Software components with defects may be embedded in end-user applications. Software developers work around known problems, making failure a rare event. We interpret this to mean that the usage distribution is adjusted to avoid a known defect. If this is achieved, the relative frequency of failure in the total usage history may approach 0.

- Growth in usage of widely-distributed, commercial software has been modeled as a power function of calendar time [1/7 1]. If the defects that cause customer-reported failures are corrected immediately for all users, the software may reach a phase in which failure is a rare event in execution intervals for a large, active group of independent users.

Thus, software which receives heavy usage according to established probabilities is a candidate for the treatment of its failures as rare events. If the likelihood of failure is indeed very small, then results in rare events and extreme values [3, 13] lead to four probability distributions as approximations under quite general conditions:

- The probability law for counts of failures in long execution intervals may be approximated as a Poisson distribution. Results of this kind are sometimes said to be a consequence of the “Poisson law of small numbers” [3]. The Poisson distribution has been introduced in other ways in the software reliability literature and already plays an important role in several software reliability models [9].

- The exponential distribution may approximate the probability law for the interoccurrence time of rare failures in long intervals. This distribution is sometimes assumed for the TTF in software reliability computations because it is implied by certain empirical studies (cf. [1]).

- Given a set of random samples of the TTF, the minimum value in the set is a random variable bounded away from 0 and its limiting distribution (for number of samples \( N \to \infty \)) is Weibull. The Weibull distribution itself has been used or suggested for other phenomena (cf. [17]).

- If the distribution of the TTF falls off fast enough for large values (see subsection 4.1), a counterpart to Weibull for the minimum TTF is Gumbel as the limiting distribution for the maximum TTF.

When conditions for rare events are met, reliability analysis in greater detail with fewer assumptions may be possible, there may be additional justification for using popular Poisson and exponential distributions, and Weibull and Gumbel distributions may also be applicable. This paper discusses aspects of rare events and extreme values in the Markov chain model
M described in the next section. Section 3 describes a Poisson Approximation for counts of a rare fail-state F in realizations of M and indicates approximation of interoccurrence times of F by an exponential distribution. Weibull and Gumbel distributions are discussed in section 4 for the extreme values of interoccurrence times of F, and section 5 illustrates computation based on the Weibull as the approximate distribution for the minimum TTF.

2 Markov Chain Model

A Markov chain can provide not only a convenient definition of software failure as a rare event (as a visit to an abnormal fail-state) but also a direct measure of rarity (using the steady-state probabilities of a recurrent chain). We will use the following model and assumptions.

Both the software usage distribution and the likelihood of failure are represented by a finite-state, discrete-parameter, time-homogeneous Markov chain M [10, 16, 24]. M has at least three states: starting-state S, terminate or end-state H, and fail-state F. State sequences are realizations of M. A realization from S to first occurrence of H represents a single successful execution cycle through the software. A transition from any state to F represents a failure of the software at that location in a realization; thus, occurrences of state F are identified with the failure events.

The probabilities on arcs in M may be actual relative frequencies from test or usage data [24] or may be values estimated for the usage distribution and component reliabilities expected in practice. Note that if no failure occurs in observed data with real software, then arcs to F with frequency count 1 are conservative assumptions in this sense: if state i, i ≠ F, has been visited n_i times and exited without failure, and if the data-based relative frequency on each arc is close to the value defined by the expected usage, then this data implies that the probability of failure at i is not greater than 1/(n_i + 1). A conditional case study could set the count on an arc from i to F to 1 to give probability p_{iF} = 1/(n_i + 1), renormalize other p_{ij} accordingly, and compute values for this hypothetical situation. If the relative frequencies p_{iF} make F a rare state, then the results in this paper are applicable.

We require that Markov chain M have (i) all states, including F, reachable from S by paths with nonzero probability, and (ii) arcs from both F and H to S with p_{FS} = p_{HS} = 1 so that both a successful termination in H and an unsuccessful run to F cause an immediate reset/restart in initial state S. This makes M a recurrent chain for studying long-term properties [10, 16]: the probability of eventually reaching every state from every state is 1. A long realization of M will ultimately include all states and, as its length increases, will become statistically typical of the entire chain in the frequencies of occurrence of all states and arcs. Let P = [p_{ij}] denote M’s transition probability matrix. M’s steady-state probability distribution \Pi = [\pi_S, \ldots, \pi_F] is the unique solution of \Pi = \Pi P where \sum_i \pi_i = 1 and \pi_i > 0 is the limiting relative frequency of occurrence of state i as a count of transitions. The mean recurrence time of state i is m_{ii} = 1/\pi_i and the mean number of occurrences of state j between visits to state i is \pi_j/\pi_i [16].

\footnote{Count of transitions is a natural reference for computation with chain M. Standard terminology uses the word time instead of the phrase count of transitions, e.g., recurrence time, interoccurrence time, time of first occurrence.}
A visit to fail-state $F$ must be a rare event, that is, $\pi_F$ must be smaller by at least several orders of magnitude than $\pi_i$ for any other state $i \neq F$. Put another way, mean recurrence time $m_{FF} = 1/\pi_F$ must be larger by orders of magnitude than $m_{ii}$ for all $i \neq F$. $m_{FF}$ is large but finite because $\pi_F$ is small but nonzero.

To focus on execution intervals between failures, we designate a realization of $M$ from $F$ to first recurrence of $F$ a run. The run-length random variable is the interoccurrence time of $F$; this is the TTF and its mean value $m_{FF}$ is the MTTF. Since $p_{FS} = 1$, runs are equivalent to realizations from starting-state $S$ to first occurrence of $F$ except for the initial transition $F$-to-$S$. The ensemble of all runs is infinite but has a well-defined, discrete probability distribution assigned by chain $M$.

3 The Poisson Approximation: Counts of Rare Fail-State $F$

Two basic and enduring number laws in probability are the well-known Gaussian Approximation which is a “Central Limit Theorem” [10] for sample averages, and the less accessible Poisson Approximation which is a “Law of Small Numbers” [3] for rare, uncommon, unlikely events. Extensive discussion of the Poisson Approximation in numerous applications is given in [3]. Po$(\lambda)$ denotes the Poisson distribution with parameter $\lambda$, that is, the discrete probability mass function \[ p(k; \lambda) = \frac{\lambda^k}{k!} e^{-\lambda} \]
for $k = 0, 1, 2, \ldots$. Po$(\lambda)$ has mean and variance equal to $\lambda$. When $k$ is the largest integer not greater than $\lambda$, $p(k; \lambda)$ has as large a value as it does for any $k$.

Visits to rare state $F$ in recurrent Markov chain $M$ are crucial events. Let random variable $n_F$ be the number of visits to $F$ in a randomly generated realization of $n$ transitions starting in state $S$. Let $\mathcal{L}(n_F)$ be the true probability law of $n_F$. Write $\lambda = E(n_F)$ for the mean value. Developments in small number laws stress that Po$(\lambda)$ is an approximation of $\mathcal{L}(n_F)$ and compute an upper bound on a measure of distance between them. The total variational distance $d_{TV}[\mathcal{L}(n_F), Po(\lambda)]$ is defined as
\[ d_{TV}[\mathcal{L}(n_F), Po(\lambda)] = \sup_A |\mathcal{L}(n_F)(A) - Po(\lambda)(A)| \]
for events (subsets) $A$ in the sample space [3]. The value is $0 \leq d_{TV} \leq 1$. With reference to $\lambda$, $\pi_F$, and $j$-step transition probabilities $p_{FF}^{(j)}$, an explicit bound derived in [3] is
\[ d_{TV}[\mathcal{L}(n_F), Po(\lambda)] \leq (1 - e^{-\lambda}) \left( \pi_F + 2 \sum_{j \geq 1} |p_{FF}^{(j)} - \pi_F| \right). \]
Since $\pi_F$ equals the limiting relative frequency of state $F$, for large $n$ \[ \pi_F \approx \frac{E(n_F)}{n} = \frac{\lambda}{n} \] and $\lambda \approx n \pi_F$. 

5
The approximation
\[ \mathcal{L}(n_F)(k) \approx \frac{(n \pi_F)^k}{k!} e^{-n \pi_F} \]
is for large \( n \) and small \( \pi_F \). \( \lambda \approx n \pi_F \) is the approximate parameter for a full sequence of \( n \) transitions, not per transition, and Poisson \( p(k; n \pi_F) \) is the approximate probability of \( k \) visits to \( F \) in a full sequence.

The interocurrence times of rare state \( F \) in \( M \) are iid (due to the Markov property) according to a discrete distribution with mean value \( m_{FF} \). The exponential probability law \( \text{Exp}(\beta) \) has a continuous cumulative distribution function (Cdf) of the form
\[ F_E(t) = \text{Prob}(T \leq t) = 1 - \exp \left\{ -\frac{t}{\beta} \right\} \]
with mean value \( \beta \) for \( T \geq 0 \). If the number of transitions \( n \) is very large relative to \( m_{FF} \), the discrete distribution for the TTF can be approximated by an exponential model as follows.

Take any long, random realization of \( n \) transitions and mark the visits to non-rare state \( S \). This partitions the realization into subsequences with mean length \( m_{SS} \). Since state \( F \) has either \( 0 \) or \( 1 \) occurrences in each subsequence, there is an independent Bernoulli trial \([1,0]\) for each subsequence with probabilities
\[ p = \text{Prob}(1 \text{ occurrence of } F), \quad q = \text{Prob}(0 \text{ occurrences of } F) = 1 - p \]
where \( p << q < 1 \) because \( F \) is rare but \( S \) is not. But the mean number of occurrences of \( F \) between occurrences of \( S \) equals \( p \), that is, for the 0-1 Bernoulli random variable
\[ \frac{\pi_F}{\pi_S} = 1 \cdot p + 0 \cdot q = p. \]

Given that \( F \) has just occurred, let \( \nu \) be the number of subsequences until its next recurrence. \( \nu \) has a geometric distribution \([10]\). Cumulative geometric probabilities \( P_\nu \) and their exponential approximations are:
\[
\begin{align*}
P_{\nu}(\nu \leq 1) &= p \approx 1 - e^{-p} = F_E(1), \\
P_{\nu}(\nu \leq 2) &= p + qp \approx 1 - e^{-2p} = F_E(2), \\
P_{\nu}(\nu \leq 3) &= p + qp + q^2 p \approx 1 - e^{-3p} = F_E(3), \\
&\vdots
\end{align*}
\]
where \( F_E(m) \) is the exponential Cdf, \( 1 - e^{-mp} \). These approximations are based on the exponential series \([18]\)
\[ e^{-mp} = 1 - mp + \frac{(mp)^2}{2!} + \cdots \]
and inequalities \( p >> p^2 >> p^3 >> \cdots \). The expected value \( E(\nu) \) for the geometric distribution and for its approximation is \( 1/p = \pi_S/\pi_F = m_{FF}/m_{SS} \), the mean number of occurrences of \( S \) between occurrences of \( F \). Since the mean subsequence length is \( m_{SS} \) transitions, the expected count of transitions between visits to \( F \) is
\[ m_{SS}E(\nu) = m_{SS} \frac{m_{FF}}{m_{SS}} = m_{FF} \]
as required. This exponential approximation for the TTF does not depend on the detailed structure of $M$ beyond the requirement for a 0-1 count of $F$'s between visits to $S$.

**Example 1:** A high level, five-state Markov chain for operation of three windows in a graphical user-interface is given in [23]. Here, the fail-state $F$ shown in figure 1 has been added as a sixth state to represent failure events associated with state $C$, e.g., as failures actually observed during testing or as an estimate of less-than-perfect reliability of a software component. The transition probabilities $p_{ij}$ for the ordinary usage-states $i \neq F$, $j \neq F$ represent the usage distribution, except that $p_{CA} = 1$ is reduced by small probability $p_{CF}$. For the range $0 < p_{CF} \leq 10^{-3}$, we compute $0 < \pi_{F} \leq 1.43 \times 10^{-4}$ and find that all other $\pi_i$'s have relatively stable values. Specifically for $p_{CF} = 10^{-4}$, the stationary distribution vector is

$$
\pi = [\pi_S, \pi_A, \pi_B, \pi_C, \pi_H, \pi_F] = [0.14278, 0.42843, 0.14283, 0.14283, 0.14276, 1.42689 \times 10^{-5}],
$$

the vector of mean recurrence times is

$$
[m_{SS}, ..., m_{FF}] = \left[ \frac{1}{\pi_S}, ..., \frac{1}{\pi_F} \right]
\quad = \left[ 7.00400, 2.33412, 7.00119, 7.00120, 7.00470, 7.00821 \times 10^4 \right],
$$

the vector of the expected number of occurrences of states between visits to non-rare state $S$ is

$$
\frac{\Pi}{\pi_S} = [1, 3.0007, 1.0004, 1.0004, 0.9999, 0.0001],
$$

and the vector of the expected number of occurrences between visits to rare state $F$ is

$$
\frac{\Pi}{\pi_F} = 10^4 \times [1.0006, 3.0025, 1.001, 1.001, 1.0005, 0.0001].
$$

For $p_{CF} = 10^{-5}$, these vectors are

$$
\Pi = [0.14277, 0.42844, 0.14284, 0.14284, 0.14276, 1.42693 \times 10^{-6}],
$$

$$
[m_{SS}, ..., m_{FF}] = [7.00445, 2.33406, 7.00102, 7.00102, 7.00452, 7.00803 \times 10^5],
$$

$$
\frac{\Pi}{\pi_S} = [1, 3.001, 1.0005, 1.0005, 0.99999, 9.9949 \times 10^{-6}],
$$

and

$$
\frac{\Pi}{\pi_F} = 10^5 \times [1.0005, 3.0025, 1.001, 1.001, 1.0005, 0.00001]
$$

respectively.

The mean recurrence time of state $S$ is $m_{SS} \approx 7$. $n = 7 \times 10^4$ transitions correspond to about $10^4$ average sequences from $S$ to $S$. We fix $n$ at $7 \times 10^4$ and evaluate the Poisson Approximation of $\mathcal{L}(n_F)$ for four values of $p_{CF}$: $10^{-4}$, $10^{-5}$, $10^{-6}$, and $10^{-7}$. 

7
lists observed frequencies

Example 2: Sample data was obtained by using random-number generator `rand` in MATLAB (R) to simulate \( \textbf{M} \) for \( 7 \times 10^4 \) transitions starting in state \( S \). Visits to state \( F \) were then counted. The experiment was repeated 1000 times for \( p_{CF} = 10^{-4} \) and for \( p_{CF} = 10^{-5} \).

For \( p_{CF} = 10^{-4} \), the 1000 samples were grouped into five classes: \( C_1 \) for count 0, \( C_2 \) for count 1, \( C_3 \) for count 2, \( C_4 \) for count 3, and \( C_5 \) for count \( \geq 4 \). Table 3 in the Appendix lists observed frequencies \( f_i \), expected frequencies \( 1000p_i \), and \( \chi^2 \)-residuals for Po(1). The \( \chi^2 \) critical value for level of significance 0.1 and 4 (=5-1) degrees of freedom is 7.78. The statistic 1.2071 is less than the critical value; therefore, in terms of testing for goodness of fit at significance level 0.1, there is no reason to reject the hypothesis that the Poisson Po(1) models the phenomenon that gave rise to the 1000 samples.

For \( p_{CF} = 10^{-5} \), we combine classes at higher counts where expected frequencies are low.
A rule-of-thumb for the $\chi^2$ test is to define the classes so that most expected counts exceed 5, but one or two can be less than 5 [6]. Classes are $C_1$ for count 0, $C_2$ for count 1, and $C_3$ for count $\geq 2$. Table 4 presents the numbers. The $\chi^2$ critical value is 4.61 for 2 (=3-1) degrees of freedom and level of significance 0.1. The statistic 1.4229 is less than 4.61, so the hypothesis is not rejected at significance level 0.1.

The $TTF$ in this software model is the interoccurrence time of rare fail-state $F$. The $\chi^2$ goodness-of-fit test can also be applied to the hypothesis that this interoccurrence time has an exponential distribution. A test can be computed in the following way without an explicit value for the parameter $\beta$ of $\text{Exp}(\beta)$ [11]. Given that $N$ visits to $F$ occur in time interval $(0,T)$, let random variables $T_1, T_2, \ldots, T_N$ be the interoccurrence times and let random variables

$$A_1 = T_1, \quad A_2 = T_1 + T_2, \quad \ldots, \quad A_N = T_1 + \cdots + T_N$$

be the waiting times. If the interoccurrence times $T_i$ are $iid$ according to an exponential distribution, the waiting times $A_i$, taken as $N$ unordered samples, are independently and uniformly distributed in $(0,T)$ [4]. Thus, the interval $(0,T)$ can be partitioned into $k$ equal subintervals, each subinterval having length $T/k$ and (due to uniformity) each having $N/k$ as the expected number of visits. Let $f_i$ be the number of visits actually observed in subinterval $i$; then the statistic

$$\sum_{i=1}^{k} \left( \frac{f_i - \frac{N}{k}}{\frac{N}{k}} \right)^2$$

is approximately $\chi^2$ with $k-1$ degrees of freedom, provided that $N \geq 20$ and each $f_i$ is sufficiently large [11]. If the statistic is less than the appropriate critical value, the hypothesis of exponential interoccurrence times is not rejected; otherwise, it is rejected.

1000 samples of interoccurrence times of rare fail-state $F$ in $M$ in figure 1 were generated for $p_{CF} = 10^{-4}$ and for $p_{CF} = 10^{-5}$. Waiting times $A_1, \ldots, A_{1000}$ were then computed. Partitioning into $k = 20$ equal subintervals gives an expected frequency of 1000$p_k = 50$ per subinterval. See table 5 in the Appendix for the data. The $\chi^2$ critical value for level of significance 0.1 and 19 (=20-1) degrees of freedom is 27.20. As shown in the table, the statistic is 23.56 for $p_{CF} = 10^{-4}$ and is 20.04 for $p_{CF} = 10^{-5}$. At significance level 0.1, the hypothesis of exponential interoccurrence times is not rejected for either value of $p_{CF}$.

4 Distributions of Extreme Values

Estimation of the $MTTF$ is included in many models of software reliability [19]. Results in extreme values concern the minimum and maximum values of the $TTF$ as random variables in their own right. Scenarios in which stochastic properties of one or both extreme values are of interest include the following.

- Evaluation of software for safety critical applications is an important issue. A software reliability model based on statistics of the extremes for safety critical systems is developed in [14].
• Suppose widely-distributed software has a large base of independent users, the users have the same usage probabilities, and failure is a rare event. Long-term, statistical characteristics of both shortest and longest TTF provide information about expected performance encountered by the ensemble of users as a group.

• Suppose a long computation executes in parallel on a network of computers as follows. There is a time requirement: the parallel computation must have uninterrupted time interval $\tau$ simultaneously at every processor in order to complete. Processors use the same software and have the same usage distribution which includes interaction with the network. Tasks at different processors are loosely-coupled and are, to a first approximation, independent. A software failure at a processor is a rare event but may be a consequence of local data or the network. Any software failure before $\tau$ forces the entire computation to restart; but due to randomness in network characteristics, a restarted computation is not likely to be identical to previous attempts, even though the same local data is processed again. A long-run characteristic of interest is the number of times on average a computation must restart due to a failure in time less than $\tau$.

4.1 Weibull and Gumbel Distributions in the Limit

Consider Markov chain $M$ again as a source of iid runs from $F$ to the first recurrence of $F$. In extreme value analysis, the distribution of these runs is called the initial distribution. Given a set of runs, let $F_{\min}$ be the minimum run-length and $F_{\max}$ the maximum run-length. $F_{\min}$ and $F_{\max}$ are called extreme value random variables.

$F_{\min}$ is nonnegative and has an initial distribution bounded away from 0 in the direction of the extreme by the minimum-length run from $F$ to $F$. $F_{\max}$ is nonnegative and has an initial distribution that is unbounded in the direction of the extreme but, by virtue of the exponential approximation in section 3, has an exponential upper bound (that is, the initial distribution is of the exponential type [6, 13]). Generic results in limiting distributions for $F_{\min}$ and $F_{\max}$ in our model are as follows [5, 6, 13].

• The probability law of $F_{\min}$ approaches a Weibull distribution $\text{Weib}(\alpha, \beta)$ with Cdf

$$F_{W}(x) = 1 - \exp \left\{ - \left( \frac{x}{\beta} \right)^{\alpha} \right\}, \alpha > 0, \beta > 0.$$  

$\alpha$ is the shape parameter: larger (smaller) $\alpha$ implies more (less) peaked density function. $\text{Weib}(1, \beta)$ with $\alpha = 1$ is in fact the exponential $\text{Exp}(\beta)$ with mean value $\beta$. $\beta$ is the scale parameter: the distribution depends on $\beta$ and $x$ through the ratio $x/\beta$.

• The probability law of $F_{\max}$ approaches a Gumbel distribution $\text{Gumb}(\mu, \rho)$ with Cdf

$$F_{G}(x) = \exp \left\{ - \exp \left\{ - \frac{x - \mu}{\rho} \right\} \right\}, \infty > \mu > -\infty, \rho > 0.$$  

$\mu$ is the location parameter: association of $x$ with $\mu$ is of the form $x - \mu$. $\rho$ is the scale parameter.

• In the special case of $N$ samples of extreme values for which the initial distribution for
Recurrences of rare state $F$ in Markov chain $M$ is exactly the exponential $\text{Exp}(m_{FF})$, then Gumb$(\mu, \rho)$ for $F_{\text{max}}$ has parameters [6]

$$\mu = m_{FF} \ln(N) \text{ and } \rho = m_{FF}$$

and Weib$(\alpha, \beta)$ for $F_{\text{min}}$ reduces to an exponential with parameters [6]

$$\alpha = 1 \text{ and } \beta = \frac{m_{FF}}{N}.$$  

### 4.2 Maximum-Likelihood Estimates from Samples

Estimation of parameters of Weibull and Gumbel Cdf’s from samples is of interest for comparisons with the theoretical limits described above. In general, finite sets of samples do not fit the asymptotic formulations perfectly. This section describes maximum-likelihood (ML) estimation and illustrates with data from Markov chain $M$ in figure 1.

Given samples $\{x_1, x_2, \ldots, x_N\}$ of $F_{\text{min}}$, ML-estimates of $\alpha$ and $\beta$ in Weib$(\alpha, \beta)$ satisfy the equations [6]

$$\frac{1}{N} \sum_{i=1}^{N} \ln x_i = \left[ \frac{\sum_{i=1}^{N} x_i^\alpha \ln x_i}{\sum_{i=1}^{N} x_i^\alpha} \right]^{-1} - \frac{1}{\alpha}$$

and

$$\hat{\beta} = \left[ \frac{1}{N} \sum_{i=1}^{N} x_i^\alpha \right]^{-\hat{\alpha}}.$$

$\hat{\alpha}$ may be obtained from the first equation by iteration, then $\hat{\beta}$ computed by the second equation. Small $N$ biases the estimate $\hat{\alpha}$ upwards [6], that is, an unbiased estimate would be smaller. An unbiasing factor is available [21] and ranges from 0.669 to 0.99 as $N$ ranges from 5 to 120. Recall that $\alpha = 1$ in a Weibull Cdf coincides with an exponential Cdf; but even when $\alpha = 1$ should be the case, an ML-estimate $\hat{\alpha}$ precisely equal to 1 is an unlikely prospect for finite samples of real data.

Given samples $\{y_1, \ldots, y_N\}$ of $F_{\text{max}}$, the ML-estimates of $\mu$ and $\rho$ in Gumb$(\mu, \rho)$ satisfy [6]

$$\hat{\rho} = \bar{y} - \left[ \sum_{i=1}^{N} y_i \exp \left\{ -\frac{y_i}{\rho} \right\} \right] \left[ \sum_{i=1}^{N} \exp \left\{ -\frac{y_i}{\rho} \right\} \right]^{-1}$$

and

$$\hat{\mu} = -\hat{\rho} \ln \left[ \frac{1}{N} \sum_{i=1}^{N} \exp \left\{ -\frac{y_i}{\rho} \right\} \right]$$

where $\bar{y} = \sum_{i=1}^{N} y_i / N$ is the sample mean. Small $N$ biases the estimate $\hat{\rho}$ downwards [6], that is, an unbiased estimate would be larger.

**Example 3:** Assume that a group of 100 users have runs iid according to Markov chain $M$ in figure 1. $M$ was simulated using RAND in MATLAB (R) to obtain a set of 100 runs, and $F_{\text{min}}$ and $F_{\text{max}}$ for the set were recorded. The experiment was repeated 64 times to obtain $N = 64$ samples of both extreme values for $p_{CF} = 10^{-4}$ and for $p_{CF} = 10^{-5}$. Details of the $\chi^2$ goodness-of-fit tests are described in the Appendix. In summary:
For $p_{CF} = 10^{-4}$ and samples of $F_{\text{min}}$, Weibull $W_1$ with ML-estimates of parameters is not rejected at level of significance 0.1, but Weibull $W_2$ based on $\text{Exp}(m_{FF})$ as the exact initial distribution is rejected at significance levels 0.1 and 0.05.

For $p_{CF} = 10^{-5}$ and $F_{\text{min}}$, neither Weibull $W_3$ with ML-estimates nor Weibull $W_4$ based on $\text{Exp}(m_{FF})$ is rejected at level of significance 0.1.

For $p_{CF} = 10^{-4}$ and $F_{\text{max}}$, Gumbel $G_1$ with ML-estimates of parameters is not rejected at level of significance 0.1, but Gumbel $G_2$ based on $\text{Exp}(m_{FF})$ is rejected at significance levels 0.1 and 0.05.

For $p_{CF} = 10^{-5}$ and $F_{\text{max}}$, Gumbel $G_3$ with ML-estimates is not rejected at level of significance 0.1. Gumbel $G_4$ based on $\text{Exp}(m_{FF})$ is rejected at level of significance 0.1 but is not rejected at reduced significance level 0.05.

None of the distributions $W_1, W_3, G_1, G_3$ with ML-estimates of parameters from samples is rejected at level of significance 0.1. Distributions $W_2, W_4, G_2, G_4$ are based on an assumption that an exponential $\text{Exp}(m_{FF})$ is exactly the initial distribution. $W_2$ and $G_2$ are obtained for state $F$ being less rare; $W_4$ and $G_4$ are obtained for state $F$ being more rare. At significance level 0.1, $W_2$ is rejected but $W_4$ is not; at significance level 0.05, $G_2$ is rejected but $G_4$ is not.

We conjecture that one factor in the outcomes of these tests is less error in the exponential approximation of the TTF's distribution for more rare $F$ than for less rare $F$.

5 Extreme Value $F_{\text{min}}$ of Software TTF

Analysis of extreme values is rich in details. We will use the Weibull $F_W$ as the approximate Cdf for $F_{\text{min}}$ to compute how often on average the minimum TTF experienced within a group of runs that are iid according to the same probability model is not less than $\tau$ where $\tau$ is a fixed reference.

An occurrence of $F_{\text{min}} > \tau$ is called an exceedance of $\tau$ [3, 6] and its complement $F_{\text{min}} \leq \tau$ is called a nonexceedance. As shown in figure 2, the binary events exceedance and nonexceedance define a Bernoulli trial [10] with probabilities

$$\text{Prob}(F_{\text{min}} > \tau) = 1 - F_W(\tau), \quad \text{Prob}(F_{\text{min}} \leq \tau) = F_W(\tau).$$

Let random variable $n_{\tau}$ be the number of independent trials from one nonexceedance of $\tau$ until the next. A standard result [10] is that $n_{\tau}$ is geometrically distributed with mean value $E_W(n_{\tau}) = 1/F_W(\tau)$. The mean number of times $\tau$ is exceeded in between nonexceedances is

$$E_W(n_{\tau}) - 1 = \frac{1}{F_W(\tau)} - 1 = \frac{1 - F_W(\tau)}{F_W(\tau)} = \frac{\text{Prob}(F_{\text{min}} > \tau)}{\text{Prob}(F_{\text{min}} \leq \tau)}.$$
Example 4: Weibulls $W_1$ and $W_3$ above are for a group of 100 runs, say, as experienced by 100 independent users with the same usage distribution. $W_1$ is obtained for $MTTF = 70082$ and $W_3$ for $MTTF$ about 10 times longer. For $\tau = 500$,

$$F_{W_1}(500) = 0.6076, \quad E_{W_1}(n_{500}) = \frac{1}{0.6076} = 1.6457$$

$$F_{W_3}(500) = 0.0325, \quad E_{W_3}(n_{500}) = 30.7630.$$ Based on $W_1$ for $F_{min}$ in the group of 100 users, on average the group experiences 0.6457 runs with minimum $TTF > 500$ in between occurrences of minimum $TTF \leq 500$. Based on $W_3$ for $F_{min}$, the comparable figure is 29.763 runs. Doubling $\tau$ to 1000 and again to 2000 gives

$$E_{W_1}(n_{1000}) = 1.1959, \quad E_{W_3}(n_{1000}) = 14.2553$$

$$E_{W_1}(n_{2000}) = 1.0312, \quad E_{W_3}(n_{2000}) = 6.76056.$$ Rather than a group of different users, $W_1$ or $W_3$ could describe $F_{min}$ for blocks of 100 runs by a single user. Suppose $W_3$ applies for a single user $U$ who starts up a block of 100 tasks every day, each of which executes $iid$ according to $M$ in figure 1 and each of which must have execution-time exceeding $\tau$. Then user $U$ will encounter a day with $F_{min} \leq \tau$ on average about once each month ($\sim 30.76$ blocks) for $\tau = 500$, about once each fortnight ($\sim 14.26$ blocks) for $\tau = 1000$, or about once each week ($\sim 6.76$ blocks) for $\tau = 2000$. If user $U$ voluntarily stops the failure-free executions after $\tau$ is reached in one day and restarts all tasks the next day, the next sample of $F_{min}$ is generated independently according to the same Cdf $F_{W_3}$ (there is no parameter-estimation with data that is censored by the non-failure stops of software execution [7]). Given that the event $F_{min} \leq \tau$ does occur, the conditional mean values of $F_{min}$ for the three values of $\tau$ are

$$E_{W_3}(F_{min} | F_{min} \leq 500) = 264.9, \quad E_{W_3}(F_{min} | F_{min} \leq 1000) = 526.5,$$

$$E_{W_3}(F_{min} | F_{min} \leq 2000) = 1038.7.$$
If nonexceedance is a rare event for a given \( \tau \), then \( F_W(\tau) = 1/E_W(n_\tau) \) is small and the exponential series [18] gives as an approximation

\[
F_W(\tau) = 1 - \exp \left\{ -\left( \frac{\tau}{\beta} \right)^\alpha \right\} \\
\approx \left( \frac{\tau}{\beta} \right)^\alpha.
\]

**Example 5:** For \( W_3 \) in the example above:

\[
F_{W_3}(1000) = 0.0701 \text{ and } \left( \frac{1000}{\beta} \right)^\alpha = 0.0727, \\
F_{W_3}(500) = 0.0325 \text{ and } \left( \frac{500}{\beta} \right)^\alpha = 0.0330, \\
F_{W_3}(100) = 0.0053 \text{ and } \left( \frac{100}{\beta} \right)^\alpha = 0.0053.
\]

Suppose user \( U \) in that example wants \( \tau \) as large as possible to average a full year of successful days (that is, an average of 365 blocks having \( F_{\text{min}} > \tau \) in between blocks having \( F_{\text{min}} \leq \tau \)). An estimate of \( \tau \) for \( E_{W_3}(n_\tau) = 366 \) is

\[
\tau \approx \beta \left( \frac{1}{366} \right)^\frac{1}{\alpha} = 55.945
\]

where, in fact, \( 1/F_{W_3}(55.945) = 366.5 \).

**Example 6:** As a final example, suppose that (i) a continuously-executing application consists of many tasks distributed on a network of processors, (ii) task executions are \( iid \) according to the same Markov chain model with rare fail-state, and (iii) all current executions in the block of parallel tasks continue until one fails, at which time the system resets and all tasks immediately restart in synchronism. Then \( W=\text{Weib}(\alpha, \beta) \), which is the limiting probability law of \( F_{\text{min}} \), also approximately describes the *system-restart interoccurrence times*. As approximations, the probability that a system-restart does not occur before \( \tau \) is 

\[
1 - F_W(\tau) = \text{Prob}(F_{\text{min}} > \tau) \text{ and the expected number of exceedances of } \tau \text{ in between nonexceedances is } E_W(n_\tau) - 1.
\]

**6 Summary and Conclusions**

Markov chains have diverse uses in software reliability computations (cf, [2, 8, 12, 24]). Rare events and statistics of the extremes are applicable to restricted, but important, aspects
of software reliability analysis (cf, [14, 15]). This paper combines results in rare events and extreme values with software reliability computations based on a finite-state, discrete-parameter, recurrent Markov chain $M$. The chain provides a convenient definition of failure as a rare event, namely, as a fail-state $F$ for which the steady-state probability $\pi_F$ is orders of magnitude smaller than $\pi_i$ for ordinary usage states $i \neq F$.

When applicable, results in rare events and extreme values may provide information about software reliability, especially in details of likely performance in the long-run. Four distributions arise naturally as approximations when software failures correspond to visits to a rare fail-state $F$ in $M$:

- Poisson for count of visits to $F$ in a long realization.
- Exponential Cdf for interoccurrence time of $F$ (the TTF).
- Weibull as limiting Cdf for minimum recurrence time of $F$.
- Gumbel as limiting Cdf for maximum recurrence time of $F$.

An upper bound on $d_{TV}$ was computed for a Poisson Approximation for Markov chain $M$ in figure 1. $\chi^2$ goodness-of-fit tests were computed for approximate distributions and random samples from chain $M$. For extreme values, the parameters for $\text{Weib}(\alpha, \beta)$ and $\text{Gumb}(\mu, \rho)$ were computed by ML-estimation from samples and also by assuming an initial exponential $\text{Exp}(m_{FF})$ for the TTF. Outcomes of these tests are compatible with the general proposition that the more rare the event, the better the approximation. Although our example used a small Markov chain to demonstrate various computations, the methods apply to chains of arbitrarily large size.

One significance of the Markov chain model is that it structurally represents the software system under test or in use in the field, and indirectly represents the software complexity. If the transition probabilities are obtained from actual frequency counts during testing, an up-to-date representation of the amount and completeness of testing is provided.

Important contributions of rare events and extreme values are the four distributions listed above as approximations in the analysis of TTF, given the basic assumptions set forth. We believe that the combination of the Markov model and these results is a powerful tool for aspects of software reliability analysis.

7 Appendix

7.1 The $\chi^2$ Goodness-of-Fit Test

The test is described in numerous references (cf. [6, 11]). $H_0$ is the hypothesis that independent samples are from a candidate distribution. A statistic is computed to measure how much the samples disagree with the hypothesis. If disagreement is “significant,” $H_0$ is rejected; otherwise, it is not.

The binary outcome of a test is solely to reject or not to reject $H_0$. The $\chi^2$ distribution provides the critical value for this binary decision. This distribution’s parameter is the
**number of degrees of freedom.** If there are \( N \) samples and \( s \) parameters for the distribution under test are estimated from the data, there are \( N - s - 1 \) degrees of freedom; if no parameters are estimated, there are \( N - 1 \) degrees of freedom. The critical value is computed for a given *level of significance*, which is the probability of rejecting \( H_0 \) when \( H_0 \) is true. A higher significance level defines a more critical test than a lower one. Significance levels 0.1, 0.05, and 0.01 are commonly used in practice.

The set of possible values of samples is partitioned into \( k \) classes \( C_1, \ldots, C_k \), where \( p_i \) is the probability of an observation in \( C_i \) according to the hypothesized distribution, \( N p_i \) is the expected number of observations in \( C_i \) in \( N \) samples, and \( f_i \) is the frequency of occurrence of actual samples in \( C_i \). Two rules-of-thumb [6] are (i) the sample size \( N \) should be greater than 20, and (ii) the classes should be defined so that most expected frequencies \( N p_i \) exceed 5, but one or two can be less than 5. The test statistic

\[
\sum_{i=1}^k \frac{(f_i - N p_i)^2}{N p_i}
\]

is approximately \( \chi^2 \) with the appropriate degrees of freedom. If the statistic is less than the appropriate \( \chi^2 \) critical value, \( H_0 \) is not rejected; otherwise, \( H_0 \) is rejected because a statistic larger than the critical value implies, at the given significance level, that the observed data are not random samples from the hypothesized model.

For reference in examples, Table 2 lists \( \chi^2 \) critical values for two levels of significance and several degrees of freedom.

<table>
<thead>
<tr>
<th>level of sig.</th>
<th>degrees of freedom</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1     2     3     4     5     6     7     8     9     10    19</td>
</tr>
<tr>
<td>0.10</td>
<td>2.71   4.61  6.25   7.78   9.24  10.65  12.02  13.36  14.68  15.99  27.20</td>
</tr>
<tr>
<td>0.05</td>
<td>3.84   5.99  7.81   9.49  11.07  12.59  14.07  15.51  16.92  18.31  30.14</td>
</tr>
</tbody>
</table>

Table 2: \( \chi^2 \) critical values (rounded values from MATLAB (R) function CHI2INV)

### 7.2 Tables for Examples 1 and 2 (Section 2)

The following tables are referenced in the \( \chi^2 \) goodness-of-fit tests in section 2.
<table>
<thead>
<tr>
<th>Class $C_i$</th>
<th>Observed frequency $f_i$</th>
<th>Expected frequency $1000p_i$</th>
<th>$\chi^2$-residual $\frac{(f_i-1000p_i)^2}{1000p_i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_1$: count 0</td>
<td>366</td>
<td>367.9</td>
<td>0.0098</td>
</tr>
<tr>
<td>$C_2$: count 1</td>
<td>379</td>
<td>367.9</td>
<td>0.3349</td>
</tr>
<tr>
<td>$C_3$: count 2</td>
<td>183</td>
<td>183.9</td>
<td>0.0044</td>
</tr>
<tr>
<td>$C_4$: count 3</td>
<td>55</td>
<td>61.3</td>
<td>0.6475</td>
</tr>
<tr>
<td>$C_5$: count $\geq 4$</td>
<td>17</td>
<td>19.0</td>
<td>0.2105</td>
</tr>
<tr>
<td>sum</td>
<td>1000</td>
<td>1000.0</td>
<td>1.2071</td>
</tr>
</tbody>
</table>

Table 3: Calculation of $\chi^2$ statistic for Poisson for $p_{CF} = 10^{-4}$

<table>
<thead>
<tr>
<th>Class $C_i$</th>
<th>Observed frequency $f_i$</th>
<th>Expected frequency $1000p_i$</th>
<th>$\chi^2$-residual $\frac{(f_i-1000p_i)^2}{1000p_i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_1$: count 0</td>
<td>895</td>
<td>904.8</td>
<td>0.1061</td>
</tr>
<tr>
<td>$C_2$: count 1</td>
<td>101</td>
<td>90.5</td>
<td>1.2182</td>
</tr>
<tr>
<td>$C_3$: count $\geq 2$</td>
<td>4</td>
<td>4.7</td>
<td>0.0985</td>
</tr>
<tr>
<td>sum</td>
<td>1000</td>
<td>1000.0</td>
<td>1.4229</td>
</tr>
</tbody>
</table>

Table 4: Calculation of $\chi^2$ statistic for Poisson for $p_{CF} = 10^{-5}$

<table>
<thead>
<tr>
<th>Subinterval</th>
<th>Observed frequency $f_i$</th>
<th>$\chi^2$-residual $\frac{(f_i-50)^2}{50}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>59/53</td>
<td>1.62/0.18</td>
</tr>
<tr>
<td>2</td>
<td>42/39</td>
<td>1.28/2.42</td>
</tr>
<tr>
<td>3</td>
<td>58/45</td>
<td>1.28/0.50</td>
</tr>
<tr>
<td>4</td>
<td>38/49</td>
<td>2.88/0.02</td>
</tr>
<tr>
<td>5</td>
<td>57/42</td>
<td>0.98/1.28</td>
</tr>
<tr>
<td>6</td>
<td>49/60</td>
<td>0.02/2.00</td>
</tr>
<tr>
<td>7</td>
<td>54/61</td>
<td>0.32/2.42</td>
</tr>
<tr>
<td>8</td>
<td>53/47</td>
<td>0.18/0.18</td>
</tr>
<tr>
<td>9</td>
<td>44/38</td>
<td>0.72/2.88</td>
</tr>
<tr>
<td>10</td>
<td>38/49</td>
<td>2.88/0.02</td>
</tr>
<tr>
<td>11</td>
<td>48/54</td>
<td>0.08/0.32</td>
</tr>
<tr>
<td>12</td>
<td>54/52</td>
<td>0.32/0.08</td>
</tr>
<tr>
<td>13</td>
<td>61/48</td>
<td>2.42/0.08</td>
</tr>
<tr>
<td>14</td>
<td>56/38</td>
<td>0.72/2.88</td>
</tr>
<tr>
<td>15</td>
<td>43/56</td>
<td>0.98/0.72</td>
</tr>
<tr>
<td>16</td>
<td>59/62</td>
<td>1.62/2.88</td>
</tr>
<tr>
<td>17</td>
<td>59/53</td>
<td>1.62/0.18</td>
</tr>
<tr>
<td>18</td>
<td>44/53</td>
<td>0.72/0.18</td>
</tr>
<tr>
<td>19</td>
<td>39/46</td>
<td>2.42/0.32</td>
</tr>
<tr>
<td>20</td>
<td>45/55</td>
<td>0.50/0.50</td>
</tr>
<tr>
<td>sum</td>
<td>1000/1000</td>
<td>23.56/20.04</td>
</tr>
</tbody>
</table>

Table 5: Calculation of $\chi^2$ statistics for Uniform Waiting Times for $p_{CF} = 10^{-4}$/$p_{CF} = 10^{-5}$
7.3 $F_{\text{min}}$ and Weibull Distributions in Example 3 (Section 4)

<table>
<thead>
<tr>
<th>Class $C_i$</th>
<th>Observed frequency $f_i$</th>
<th>Expected frequency $64 p_i$</th>
<th>$\chi^2$-residual $^{[f_i - 64 p_i]^2 \over 64 p_i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_1$: $\leq 100$</td>
<td>11</td>
<td>11.7253/5.5856</td>
<td>0.0449/5.2484</td>
</tr>
<tr>
<td>$C_2$: 100* to 300</td>
<td>14</td>
<td>16.2790/9.7514</td>
<td>0.3191/1.8511</td>
</tr>
<tr>
<td>$C_3$: 300* to 500</td>
<td>12</td>
<td>10.8846/8.1235</td>
<td>0.1143/1.8498</td>
</tr>
<tr>
<td>$C_4$: 500* to 700</td>
<td>12</td>
<td>7.4646/6.7674</td>
<td>2.7509/4.0459</td>
</tr>
<tr>
<td>$C_5$: 700* to 1100</td>
<td>6</td>
<td>8.8135/10.3343</td>
<td>0.8981/1.8178</td>
</tr>
<tr>
<td>$C_6$: $&gt; 1100$</td>
<td>9</td>
<td>8.8300/23.4377</td>
<td>0.0033/8.8937</td>
</tr>
<tr>
<td>sum</td>
<td>64</td>
<td>64.0000/64.0000</td>
<td>4.1353/23.7068</td>
</tr>
</tbody>
</table>

Table 6: Calculations of $\chi^2$ statistics for Weibulls $W_1/W_2$ for $p_{CF} = 10^{-4}$

<table>
<thead>
<tr>
<th>Class $C_i$</th>
<th>Observed frequency $f_i$</th>
<th>Expected frequency $64 p_i$</th>
<th>$\chi^2$-residual $^{[f_i - 64 p_i]^2 \over 64 p_i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_1$: $\leq 2 \times 10^3$</td>
<td>7</td>
<td>9.4663/10.6841</td>
<td>0.6426/1.2704</td>
</tr>
<tr>
<td>$C_2$: $(2*4) \times 10^3$</td>
<td>13</td>
<td>9.5367/8.9005</td>
<td>1.2577/1.8882</td>
</tr>
<tr>
<td>$C_3$: $(4*6) \times 10^3$</td>
<td>7</td>
<td>8.3981/7.4146</td>
<td>0.2328/0.0232</td>
</tr>
<tr>
<td>$C_4$: $(6*8) \times 10^3$</td>
<td>11</td>
<td>7.1243/6.1769</td>
<td>2.1084/3.7660</td>
</tr>
<tr>
<td>$C_5$: $(8*10) \times 10^3$</td>
<td>6</td>
<td>5.9191/5.1457</td>
<td>0.0011/0.1418</td>
</tr>
<tr>
<td>$C_6$: $(1*1.4) \times 10^4$</td>
<td>4</td>
<td>8.7797/7.8577</td>
<td>2.6021/1.8939</td>
</tr>
<tr>
<td>$C_7$: $&gt; 1.4 \times 10^4$</td>
<td>16</td>
<td>14.7759/17.8205</td>
<td>0.1014/0.1860</td>
</tr>
<tr>
<td>sum</td>
<td>64</td>
<td>64.0000/64.0000</td>
<td>6.9461/9.1695</td>
</tr>
</tbody>
</table>

Table 7: Calculations of $\chi^2$ statistics for Weibulls $W_3/W_4$ for $p_{CF} = 10^{-5}$

For $p_{CF} = 10^{-4}$, the 64 samples of $F_{\text{min}}$ have range $[5,2912]$, mean 548.5, and standard deviation 545.05. Iteration with these samples gives $\hat{\alpha} = 0.9708$ as ML-estimate of the Weibull parameter. The unbiasing factor for 64 samples is 0.98 [21], so the unbiased ML-estimate is $\hat{\alpha} = 0.9513$. We obtain $\hat{\beta} = 536.26$ and the distribution $W_1=\text{Weib}(\hat{\alpha},\hat{\beta})$ with mean 548.45 and standard deviation 576.74.

Using $\text{Exp}(m_{FF})$ as the initial distribution for interoccurrence time of rare fail-state $F$ gives $\alpha = 1$, $\beta = 1095$, and $W_2=\text{Weib}(1,1095) = \text{Exp}(1095)$ with mean value about twice the sample mean.

$\chi^2$ goodness-of-fit tests were applied to the samples of $F_{\text{min}}$ and the separate Weibulls $W_1/W_2$. (This is not a test of one distribution vs. another as alternate hypotheses.) Samples were grouped into six classes by value. See Table 6. Since two parameters are estimated from the data for $W_1$, there are 3 (=6-2-1) degrees of freedom. Consider level of significance 0.1. The $\chi^2$ critical value 6.25 exceeds the statistic 4.1353 for $W_1$; therefore, at significance level 0.1, there is no reason to reject the hypothesis that $W_1$ models the phenomenon that gave rise to the samples. Since no parameters are estimated from the data for $W_2$, there are 5 (=6-1) degrees of freedom; the critical value 9.24 is less than the statistic 23.7068 for
$W_2$, so $W_2$ is rejected. In fact, $W_2$ is also rejected at the reduced level of significance 0.05 because the $\chi^2$ critical value 11.07 is less than its statistic 23.7068.

For $p_{CF} = 10^{-5}$, the 64 samples of $F_{\min}$ have range [23,35712], mean $9559.9$, and standard deviation 8301.7. ML-estimate is $\hat{\alpha} = 1.1613$, adjusted to $1.1381$ by the unbiasing factor and giving $\beta = 10004$. $W_3=\text{Weib}(1.1381,10004)$ has mean $9550.3$ and standard deviation 8410.4. Assuming initial distribution $\text{Exp}(m_{pF})$ gives $\alpha = 1$, $\beta = 10950$, and $W_4=\text{Exp}(10950)$ with mean value fairly close to the sample mean.

The 64 samples were grouped into seven classes. See Table 7. Consider level of significance 0.1. The critical values are 7.78 for 4 (=7-2-1) degrees of freedom and 10.65 for 6 (=7-1) degrees of freedom. Critical value 7.78 exceeds $W_3$’s statistic 6.9461 and critical value 10.65 exceeds $W_4$’s statistic 9.1695. Neither $W_3$ nor $W_4$ is rejected.

### 7.4 $F_{\max}$ and Gumbel Distributions in Example 3 (Section 4)

<table>
<thead>
<tr>
<th>Class $C_i$</th>
<th>Observed frequency $f_i$</th>
<th>Expected frequency $64p_i$</th>
<th>$\chi^2$-residual $\frac{(f_i-64p_i)^2}{64p_i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_1$: $\leq 2.7 \times 10^5$</td>
<td>5</td>
<td>6.2966/16.4534</td>
<td>0.2670/7.9728</td>
</tr>
<tr>
<td>$C_2$: $(2.7^+ to 3) \times 10^5$</td>
<td>7</td>
<td>6.3376/9.9517</td>
<td>0.0692/0.8755</td>
</tr>
<tr>
<td>$C_3$: $(3^+ to 3.3) \times 10^5$</td>
<td>8</td>
<td>7.9324/9.5351</td>
<td>0.0006/0.2471</td>
</tr>
<tr>
<td>$C_4$: $(3.3^+ to 3.6) \times 10^5$</td>
<td>11</td>
<td>8.3552/7.9985</td>
<td>0.8372/1.1263</td>
</tr>
<tr>
<td>$C_5$: $(3.6^+ to 3.9) \times 10^5$</td>
<td>8</td>
<td>7.7914/6.1483</td>
<td>0.0056/0.5577</td>
</tr>
<tr>
<td>$C_6$: $(3.9^+ to 4.2) \times 10^5$</td>
<td>7</td>
<td>6.6685/4.4631</td>
<td>0.0165/1.4420</td>
</tr>
<tr>
<td>$C_7$: $(4.2^+ to 4.8) \times 10^5$</td>
<td>5</td>
<td>9.5246/5.2503</td>
<td>2.1474/0.0119</td>
</tr>
<tr>
<td>$C_8$: $&gt; 4.8 \times 10^5$</td>
<td>13</td>
<td>11.0937/4.1993</td>
<td>0.3275/18.4441</td>
</tr>
<tr>
<td>sum</td>
<td>64</td>
<td>64.0000/64.0000</td>
<td>3.6730/30.6775</td>
</tr>
</tbody>
</table>

Table 8: Calculations of $\chi^2$ statistics for Gumbels $G_1$/$G_2$ for $p_{CF} = 10^{-4}$

<table>
<thead>
<tr>
<th>Class $C_i$</th>
<th>Observed frequency $f_i$</th>
<th>Expected frequency $64p_i$</th>
<th>$\chi^2$-residual $\frac{(f_i-64p_i)^2}{64p_i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_1$: $\leq 2.5 \times 10^6$</td>
<td>4</td>
<td>3.1986/10.5076</td>
<td>0.2008/4.0303</td>
</tr>
<tr>
<td>$C_2$: $(2.5^+ to 3) \times 10^6$</td>
<td>13</td>
<td>12.6651/15.9003</td>
<td>0.0089/0.5290</td>
</tr>
<tr>
<td>$C_3$: $(3^+ to 3.5) \times 10^6$</td>
<td>15</td>
<td>17.5685/15.0706</td>
<td>0.3755/0.0003</td>
</tr>
<tr>
<td>$C_4$: $(3.5^+ to 4) \times 10^6$</td>
<td>18</td>
<td>13.8709/10.2697</td>
<td>1.2292/5.8188</td>
</tr>
<tr>
<td>$C_5$: $(4^+ to 4.5) \times 10^6$</td>
<td>6</td>
<td>8.2948/5.9238</td>
<td>0.6349/0.0010</td>
</tr>
<tr>
<td>$C_6$: $&gt; 4.5 \times 10^6$</td>
<td>8</td>
<td>8.4021/6.3378</td>
<td>0.0192/0.4419</td>
</tr>
<tr>
<td>sum</td>
<td>64</td>
<td>64.0000/64.0000</td>
<td>2.4684/10.8214</td>
</tr>
</tbody>
</table>

Table 9: Calculations of $\chi^2$ statistics for Gumbels $G_3$/$G_4$ for $p_{CF} = 10^{-5}$

For $p_{CF} = 10^{-4}$, the 64 samples of $F_{\max}$ have range [203311.667105], mean $3.896 \times 10^5$, and standard deviation $1.063 \times 10^5$. ML-estimates for parameters of the Gumbel distribution are $\hat{\lambda} = 84002.79$ and $\hat{\mu} = 3.407 \times 10^5$. $G_1=\text{Gumb} (\hat{\mu}, \hat{\lambda})$ has mean $3.891 \times 10^5$ and standard
deviation $1.077 \times 10^6$. Computing parameters as if $\text{Exp}(m_{FF})$ were the initial distribution for interoccurrence time for rare fail-state $F$ gives $\rho = 70082$ and $\mu = 2.9146 \times 10^5$, in which case $G_2 = \text{Gumb}(\mu, \rho)$ has mean $3.3192 \times 10^6$ and standard deviation $8.9884 \times 10^4$.

The samples were grouped into eight classes by value. See Table 8. Consider significance level 0.1. $G_1$ has ML-estimation of two parameters from the sample data. The critical value for 5 (= 8-2-1) degrees of freedom is 9.24. $\chi^2$ statistic 3.6730 for $G_1$ is less than 9.24, so $G_1$ is not rejected. The second Gumbel $G_2$ is based on an assumption that the initial distribution is exactly the exponential $\text{Exp}(m_{FF})$. No parameters are estimated from the data. The critical value 12.02, for 7 (= 8-1) degrees of freedom, is less than the $G_2$ statistic 30.6775; therefore, statistical evidence rejects $G_2$ at significance level 0.1. $G_2$ is also rejected at the reduced level of significance 0.05.

For $p_{CF} = 10^{-5}$, the 64 samples have range $[2244658,5950128]$, mean $3.585 \times 10^6$, and standard deviation $7.917 \times 10^5$. ML-estimates yield $G_3 = \text{Gumb}(3.218 \times 10^6, 653981.33)$ which has mean $3.595 \times 10^6$ and standard deviation $8.388 \times 10^5$. An initial exponential $\text{Exp}(m_{FF})$ yields $G_3 = \text{Gumb}(2.9146 \times 10^6, 700800)$ which has mean $3.3191 \times 10^6$ and standard deviation $8.9881 \times 10^5$.

The samples were grouped into six classes. See Table 9. Consider significance level 0.1. The critical values are 6.25 for 3 (= 6-2-1), and 9.24 for 5 (= 6-1), degrees of freedom. $G_3$ is not rejected but $G_4$ is rejected; however, $G_4$ is not rejected at the reduced level of significance 0.05 because its statistic 10.8214 is less than the critical value 11.07.

**References**


