A SURVEY ON THE INTERVAL AVAILABILITY DISTRIBUTION OF FAILURE PRONE SYSTEMS

M. A. J. SMITH\textsuperscript{1,4}, T. AVEN\textsuperscript{2}, R. DEKKER\textsuperscript{3}
&
F. A. VAN DER DUYN SCHOUTEN\textsuperscript{4}

\textsuperscript{1} University of Salford, Department of Mathematics and Computer Science, Lancaster MS 4WT, United Kingdom.
\textsuperscript{2} Stavanger University Center, P.O. Box 2557, Ullandhaug, 4004 Stavanger, Norway.
\textsuperscript{3} Erasmus University Rotterdam, Department of Econometrics and Operations Research, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands.
\textsuperscript{4} Tilburg University, Center for Economic Research, P.O. Box 90153, 5000 LE Tilburg, The Netherlands.

ABSTRACT

In this paper a literature survey is presented on the distribution of the interval-availability of failure prone systems. The interval availability distribution is an increasingly important performance measure. It allows the evaluation of contracts in which a certain level of availability or reliability is guaranteed over a finite period of time. Emphasis is on the various approaches and approximations to find analytical expressions for the interval availability distribution.

KEYWORDS


1. INTRODUCTION

In studying the performance of a failure prone system, steady state measures like the long run availability of the system, do not always provide sufficient information for practical use.

In the oil industry the amount of oil to be delivered over a period to a client is often contractually guaranteed [1]. The information that a gas production platform is available for 360 days per year on average is not sufficient to determine appropriate penalty clauses. Short interruptions of several minutes can easily be covered by inventory, whereas a loss of production for several days may cause problems in meeting the sales contract [78, 20, 83]. In this respect the interval availability distribution is a more appropriate performance measure.

Also in the computer market, vendors have announced computer systems with a guaranteed level of availability of 95 to 100 percent in a given warranty period [54]. Substantial penalties are incurred if the

\begin{footnote}
\textsuperscript{4} The first author wishes to express his thanks to the Dutch Organisation for Scientific Research (NWO) for financial support.
\end{footnote}
level is not met. Consumers should be able to evaluate warranties because the penalties, though substantial, may not fully compensate inadequate service. Hence the distribution of the interval availability is a measure of interest to both producers and consumers.

Also we see that practically all U.S. government (DOD and NASA) contracts contain reliability clauses and specifications [41]. Again, the long term average availability does not provide the information to evaluate the contracts properly.

In this survey we highlight the different approaches and approximations to find analytical expressions for the interval availability distribution of failure prone systems. In section 2 we define the interval availability distribution. We show that from the interval availability distribution one can also derive the mission reliability and the point availability. In section 3 we consider the interval availability of a two state system and in section 4 of multi state systems. In section 5 we draw conclusions.

2. THE INTERVAL AVAILABILITY OF FAILURE PRONE SYSTEMS

Let $X(s)$ denote the state of a failure prone system at time $s$, $s \geq 0$. We assume that $X(s)$ can only take the values $0$ (down) and $1$ (up), while in section 3, $X(s)$ can take on any of a finite number of values $\{0, \ldots, n\}$. We split the state space $\{0, \ldots, n\}$ into two disjoint sets $U$ (up-states) and $D$ (down-states). We consider the random variable

$$Z_t = \int_0^t I_{\{X(s) \in U\}} \, ds,$$

where $I_{\{\cdot\}}$ denotes the indicator function. The random variable $Z_t$ represents the total operational time and $\frac{1}{t} Z_t$ is referred to as the interval availability. The interval availability distribution is denoted by

$$\Omega(x, t) = P(Z_t \leq x)$$

From the interval availability distribution one can also derive the mission reliability and the point availability.

1. The mission reliability, i.e. the probability that a system survives a complete interval $[0, t]$, is given by

$$P(Z_t = t) = 1 - \Omega(t^-, t).$$

where $\Omega(t^-, \cdot)$ denotes $\lim_{s \uparrow t} \Omega(s, \cdot)$.

2. The point or instantaneous availability, which represents the probability that the system is operational at time $t$ satisfies the following equation

$$P(X(t) = 1) = \frac{d}{dt} \int_0^t \frac{1}{t} P(X(s) = 1) \, ds = \frac{d}{dt} \int_0^t \frac{1}{t} \Omega(X(s) = 1) \, ds$$

$$= \frac{d}{dt} \int_0^t \frac{1}{t} \Omega(x, t) \, dx$$

$$= \frac{d}{dt} \int_0^t 1 - \Omega(x, t) \, dx$$

$$= 1 - \Omega(t^-, t) - \int_0^t \frac{d\Omega(x, t)}{dt} \, dx$$

provided that $\frac{d\Omega(x, t)}{dt}$ exists for $0 < s < t$. 

3. THE TWO STATE SYSTEM

The simplest system is a two state (single unit) system. Here the system comprises of one unit whose i-th life time, $\mathcal{F}_i$, has finite mean $\mu$ and variance $\sigma^2_{\mu}$ and i-th repair time, $\mathcal{R}_i$, has finite mean $\rho$ and variance $\sigma^2_{\rho}$.

The subsequent life- and repair times are assumed to be stochastically independent. In this case \( \{X(s), s \geq 0\} \) is an alternating renewal process.

In 1957 Takács [74] derived an exact expression of the interval availability distribution over \([0, t]\) assuming that time 0 coincides with the start of an up-period:

\[
\Omega(x, t) = \sum_{i=1}^{\infty} F^{(i)}(x)[R^{(i-1)}(t - x)] - R^{(i)}(t - x)
\]

\[= 1 - \sum_{i=0}^{\infty} R^{(i)}(t - x)(F^{(i)}(x) - F^{(i+1)}(x)) \quad 0 \leq x < t \tag{1}\]

with $F(t) = P(\mathcal{F} \leq t)$ and $R(t) = P(\mathcal{R} < t)$. Accordingly $F'(t)$ is continuous on the right and $R'(t)$ on the left. The superscript \((i)\) denotes the i-fold Stieltjes convolution and $F^{(i)}(t) = R^{(i)}(t) = 1, i \geq 0$.

The formula of Takács is easily adapted to situations where $t = 0$ coincides with the start of a repair or where at $t = 0$ the unit is in operation or repair for a given amount of time.

The distribution of the interval availability for a two state single unit system has been reconsidered by many authors [50, 71, 29, 28, 4, 24, 36], using different methods and assumptions.

In [74, 55, 65, 46, 49] some general characteristics and the limiting behavior of a two state system is studied. It has been shown by Takács and Rényi [74, 55], that, as a direct result of the Central Limit Theorem, the interval availability distribution approaches the Normal distribution when the length of the interval increases. Moreover, they derived expressions for the leading term in the asymptotic expansion of the mean and the variance of the interval availability.

In [65] the exact mean and variance and the constants in the asymptotic expansion of the mean and variance are given. Furthermore, it is shown that a (scaled) Beta distribution fits the interval availability distribution better than the normal distribution.

In [46, 49] limiting results are derived for wider classes of point processes than the alternating renewal process.

The formula of Takács (1) is in general not readily amenable for computational purposes. One problem is the evaluations of the convolutions, $F^{(i)}(x)$ and $R^{(i)}(x)$, which are in general not analytically computable. A solution suggested in [79] is to replace the life and repair time distributions by suitable phase type distributions yielding analytical expressions for the convolutions.

Another problem is the infinite summation. In [85] bounds are given that can be used to truncate the infinite summation properly. In case the life or repair times are exponentially distributed, simple bounds are derived in [36]. In case of exponentially distributed life times we have:

\[
e^{-(t-x)/\mu}[1 + \frac{t-x}{\mu} R(x)] \leq \Omega(x, t) \leq e^{-(t-x)(1-R(x))/\mu}
\]

Furthermore, another (new) analogy can be made. If the life or repair times are exponentially distributed then the interval availability distribution can be written as a function of a compound Poisson distribution which can be expressed as the solution of a one dimensional integral equation. A compound Poisson density reflects the probability that within some time, say \((t - x)\), the total aggregated amount
occurring at jumps of the Poisson process equals a given value \( x \). If we identify the aggregated amount of jump height with repair time then the compound Poisson density reflects the probability that the total repair time equals \( x \) in some time \( t - x + z \) and at time \( t \) the system is up. Therefore, if we let \( \omega(x, t) = d\Pi(x, t)/dx \) and \( r(x) = dR(x)/dx \) then we have

\[
\omega(x, t) = \gamma(x, t) - \int_0^x \frac{1}{\mu} \gamma(x - z, t - z)(1 - R(z)) dz
\]

Here \( \gamma(x, t) \) is the compound Poisson density which satisfies the integral equation\(^6\):

\[
\gamma(x, t) = -x \mu e^{-x/\mu} r(x) + \int_0^{x - y} \frac{xy}{(t - z)\mu} r(y) \gamma(x, t - y) dy
\]

This equation can be solved numerically by a fast and accurate method proposed in [21] or by Panjer's recursion [55, 22], which yields for discrete-(ized) repair time distributions, \( r_i = P(\text{repair time} = i) \), the recursive relation

\[
\gamma(x, x + n) = \sum_{j=1}^n \frac{x + n}{n\mu} (x + n - j) \gamma(x, x + n - j) \quad \text{and} \quad \gamma(x, x) = e^{x/\mu}
\]

In case \( \rho < \mu \), \( \gamma(x, t) \) is a good approximation of \( \omega(x, t) \) with \( \int_0^x \gamma(s, t) ds < O(x, t) \).

4. MULTISTATE SYSTEM

Multistate systems can be classified into two groups: Markovian and Non-Markovian. Consider the situation where \( X(s) \) can take any state from the finite set of values \( \{0, \ldots, n\} \). The state space \( S = \{0, \ldots, n\} \) is divided into two groups \( U \) (up-states) and \( D \) (down-states).

4.1. Markovian and Semi-Markovian Environment

Consider, for example, a multi-component system with \( n \) different units with exponentially distributed individual life and repair times. For this system \( X(t) \) results in a Markov chain on a state space with \( |S| = 2^n \). The subsets \( U \) and \( D \) are determined by the structure function of the system. In such a case an exact computation of the interval availability distribution is possible. Mainly the computer oriented journals there has been a great interest in exploring this direction, cf. [54, 19, 73, 75, 34, 69, 48, 33, 23, 18, 57, 56, 61, 62, 13]. Basically the following methodology is adapted:

Let \( \{X(t), t \geq 0\} \) be a continuous-time Markov process on the state space \( S = \{1, \ldots, n\} \) with infinitesimal generator \( Q \). Let \( P \) denote the one step transition matrix of the embedded Markov chain \( \{X_n, n \geq 0\} \) where \( X_n := \text{the state of the system just after the n-th transition epoch} \). By applying the well-known uniformisation technique we have for the transient probability distribution \( \pi_t \)

\[
\pi_t = \sum_{i=0}^{\infty} \frac{e^{-\lambda t}}{i!} \lambda^i P_i(\{q\})
\]

where \( q \) is the largest element of \( Q \) and \( P = Q/q + I \). Let \( \pi \) denote the steady state distribution. In [37] the approximation \( \tilde{\pi} = \sum_{i=0}^K e^{-\lambda t} \frac{\lambda^i}{i!} \left[P_i - \pi\right] \) for \( \pi_t \) is given, where the truncation level \( K \) can be chosen such that the approximation has a specified level of accuracy uniformly in \( t \).

\(^6\)This is a result from insurance theory. The integral equation is used to compute the aggregate claim size distribution. Claims arrive according to a Poisson process with rate \( 1/\mu \) and the claim amount is distributed according to \( R(x) \). The probability that the aggregate claim size in a period of length \( t \) equals \( x \) is equal to the probability that the total down time equals \( x \) in a period of length \( t + x \) and that at time \( t + x \) the system is up.
Let $W(n, k)$ denote the probability that the embedded Markov chain has visited the set $\mathcal{U}$ $k$ times in $n$ transitions. This probability can be computed recursively. The interval availability distribution can now be obtained by noticing that given $n$ events in a time interval $t$ generated by a Poisson process and given that the discrete Markov chain visited $\mathcal{U}$ $k$ times, the probability that the total time spent in $\mathcal{U}$ is smaller than or equal to some value $x$ equals the probability that the $k$-th order statistic from $n$ uniformly distributed random variables is smaller than $x/t$. Hence we have (see [20, 19, 18]):

$$
\Omega(x, t) = \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \sum_{k=0}^{n} W(n, k) \sum_{i=0}^{\frac{n}{k}} \binom{n}{i} \left( \frac{x}{t} \right)^i \left( 1 - \frac{x}{t} \right)^{n-i}
$$

The major drawback of using Markov theory is that two problems arise, namely largeness and stiffness. Largeness reflects the fact that adding e.g. a component or an exponential phase in the life time distribution, doubles the state space. Stiffness is caused by the simultaneous occurrence of small (life times) and relatively large (repair times) transition rates, causing the transition matrix to become unstable. For some interesting articles that address these problems we refer to [32, 5, 59, 60].

The extension to a semi-Markovian environment is treated by Csenki [12, 15, 17, 16, 14, 11, 10] and Rubino and Sericola [56]. In these papers the interval availability distribution is obtained using renewal theory and is given by the solution of a system of two dimensional integral equations. In [12, 15, 17, 16, 14, 11] attention has been paid to solving these equations numerically. The two dimensional integral equation is approximated using the two point trapezoidal rule in two dimensions and discretizing $\Omega(x, t)$ in the points $x = i\Delta$ and $t = j\Delta$. Next a recursion scheme for $\Omega(i\Delta, j\Delta)$ is obtained, which is solved in $O(\Delta^{-2})$ time and has an accuracy of $O(\Delta^2)$.

4.2. Non-Markovian systems

In most practical situations the system under consideration can not be modelled as a (semi) Markov model or solving the (semi) Markov model is found to be too time consuming. Then, a useful approach is to approximate the interval availability distribution by Takács formula for two state single unit systems where the up and down time distributions are given by the stationary up and down time distribution, see e.g. [78, 83, 82]. That is, let $U_s (D_s)$ be the length of the first up (down) period if at time $s$ the system just became up (down). Then the stationary up (down) time distribution, $U(t)$ ($D(t)$), is given by

$$
U(t) = \lim_{s \to \infty} P(U_s \leq t) \quad \text{and} \quad D(t) = \lim_{s \to \infty} P(D_s \leq t)
$$

In [83, 82] the following methodology is proposed (called STAMP : State space Aggregation, Markov chain Analyses and Phase type distributions).

a. Approximate all life and repair time distributions by phase type distributions.

b. With all the distributions being of phase-type it becomes possible to describe the system by a large Markov model.

c. Compute the first two moments of the stationary sojourn time in $\mathcal{U}$ and $\mathcal{D}$.

d. Fit a phase type distribution to those moments to obtain an approximation of the stationary up and down time distributions.

e. Apply the formula of Takács.
This methodology is used in [79] to compute the interval availability distribution of a $k$ out of $n$ system, with ample repair facility. In [78] the methodology is used to deal with a two-unit standby system with Markovian degrading units and generally distributed preventive and corrective repair time distributions. In this case it is possible to omit the steps $a$ and $b$.

The advantage of this approximative method over exact computation in a Markov environment is that the probabilities $W(n, k)$ need not be computed. However, also this method suffers from largeness and stiffness.

4.2.1. Regenerative systems

There is an abundant literature in which a renewal theoretical analysis is used on regenerative systems. In the 1985 literature review given by Yearout et al. [86] concerning standby redundancy 156 articles are considered. In each article the system is such that suitable regeneration points can be identified. More references can be found in [44, 52]. However, it must be noted that in almost all these articles only the long run average availability is considered. Exceptions are [72, 81, 27].

In [72] a general analysis and a short overview is given of two unit warm standby systems with preventive maintenance. Emphasis is on the computation of the first moments of the up and down time distribution. This gives us some information about the interval availability distribution.

In [81] a recursive algorithm is presented to compute all the moments of the stationary up time distribution in a 1 out of $n$ system assuming exponential repair time distributions. For a $k$ out of $n$ system with general distributions and a single repair facility the availability and stationary down time distribution has been considered in [27].

4.2.2. Non-regenerative systems

Usually one of the assumptions underlying the models in literature is that some distributions are exponential. This assumption guarantees a regenerative system. For electronic equipment or equipment that consists of many renewable non-critical parts, it is usually very well justified to use the exponential distribution for the life time distribution of a component, see e.g. [9, 25, 47, 51, 87]. However, when equipment wears out the life time distribution is usually better described by a Weibull distribution, cf. [3, 6, 40, 45] and in many situations, the repair times are best described by the log-normal distribution, cf. [8, 31, 38, 63, 84]. When there are no assumptions concerning exponential distributions there usually exist no regenerative points or the distribution of the time between two regeneration points can not be obtained. In this case one has to resort to approximations.

In [80] an approximation is given of the mean stationary up and down time of a 1 out of 2 system with cold standby, general life and repair time distributions and ample repair facility.

Based on [80], in [68] the first three moments of the stationary up and down time distributions of a 1 out of $n$ system with cold standby and general life and repair time distributions are approximated. In [64] the mean up and down times are approximated for a $k$ out of $n$ system with cold standby, general life and repair time distributions and ample repair facility. Moreover, it is explained in [66] that in case of a constant failure rate, the first moment of the stationary up and down time distribution are insensitive to the shape of the repair time distribution.

In [36] an approximation is given of the down time distribution in a 1 out of $n$ system with hot standby, ample repair facility and exponential life time distributions. For series and $k$ out of $n$ systems
with hot standby, assuming general life and repair time distributions and ample repair facility an approximation of the stationary up and down time distribution is given in [67]. Again we note that all these approximations primarily concern the steady state up and down time distributions. However, as we explained before the moments of these distributions can be used in finding approximations for the interval availability distribution.

4.2.3. Asymptotic methods

Sometimes, when the system has a high level of availability, one can make use of asymptotic methods. Let us assume that during an up period the system regenerates itself a number of times and during a regeneration period a system failure can occur with a very small probability $p$. That is, let $U$ be the length of an uninterrupted up period, $S_i$ the length of time between the $i$-th and $(i-1)$-th renewal, $L$ the number of consecutive regenerative cycles without failure and $\eta \leq S_L$, the length of the up time in the cycle with the first failure. Then

$$U = \sum_{i=1}^{L-1} S_i + \eta \text{ and } P(L = l) = (1 - p)^{l-1}p$$

Let $\mu := E(S_i)$ and $c^2$ the coefficient of variation of $S_i$. In [2] it is proved that if $F$ and $R$ varies such that $p \to 0$ and $pc^2 \to 0$, the distribution of the up time tends to an exponential distribution, i.e.

$$\lim P(U \leq pt/\mu) = 1 - \exp(-t)$$

Similar results can be found in e.g. [77, 76, 70, 30, 35, 43, 42, 39]. Additionally, in [2] it is proved that the interval availability distribution tends to a compound Poisson distribution, where the rate of the underlying Poisson process equals $p/\mu$. In [2] it has also been shown that $p/\mu$ tends to the average system failure, which is much easier to compute. In [70, 26] an estimation of the probability $p$ is given for some general renewable systems. In [39] the probability $p$ is estimated by simulation for a 1 out of $n$ system with general life and repair time distributions. In [68] the first three moments of the up and down time distributions of the 1 out of $n$ system are approximated and it is shown that the convergence of the coefficient of variation to one is rather slow and not monotonic, i.e. it increases first before it decreases to one. In [7], Brouwers computes asymptotic expressions for the mean up and down time of series, parallel, and $k$ out of $n$ systems with cold standby assuming exponential life and repair time distribution, ample repair facility by means of Markov theory. Subsequently, the asymptotic result is used to approximate the up time distributions by exponentials.

5. CONCLUSION

The interval availability distribution is an increasingly important performance measure for both producers and consumers. It is also a difficult performance measure to obtain. Only for two state (single unit) systems explicit expressions can be obtained, which are not always amenable for numerical evaluation. For multi-state system exact expressions can only be obtained in the Markovian and semi-Markovian case. However, it should be noted that the methods can be quite time and memory consuming.

In general, it seems more practical to aggregate a multi state system into a two state system. In this situation one needs to compute or approximate the stationary up and down time distributions of the aggregated system. A general methodology to do this is STAMP which makes use of Markov theory
and Phase type distributions. However, this methodology also has the drawback that the state space easily explodes.

A considerable amount of literature has appeared in which regenerative systems are analysed. However, most papers only deal with long run performance measures.

The literature in which non-regenerative models are considered is quite limited. The reason probably is that, in general, these systems can not be analysed exactly and approximations need to be made.

The asymptotic methods are very powerful for approximating the stationary up time distribution. For regenerative systems the up time distribution tends to an exponential distribution as the availability increases. For many non-regenerative systems, for which the stationary distributions exist, it is likely that this result can be extended. However, this is a topic of further research.

References


Advances in Safety and Reliability: ESREL '97


