Special Proceedings: ISBN #: 978-81-950383-0-5
23rd Annual Conference, 24-28 February 2021; pp 15-29

# Cost Minimization under Sporadic Shocks and Healing Impetus when the Healing Stage is Subdivided

**Debolina Chatterjee and Jyotirmoy Sarkar**
*Department of Mathematical Sciences*
*Indiana University-Purdue University Indianapolis, Indiana, USA*

Received: 18 May 2021; Revised: 28 May 2021; Accepted: 31 May 2021


## Abstract

A system experiences two kinds of sporadic impacts: valid shocks (VS) that cause damage, and positive interventions (PI) that induce partial healing. The system lifetime is divided into two stages: In Stage 1, healing can occur; but in Stage 2, no healing is possible. Stage 1 is further subdivided into two parts: In the early part, called Stage 1A, healing happens faster than in the later stage, called Stage 1B. The system stays in Stage 1A until the net count of impacts (VS registered *minus* VS nullified) reaches a predetermined threshold $m_A$; then the system enters Stage 1B and stays there until the net count reaches another predetermined threshold $m_1$ ($> m_A$). Thereafter, the system enters Stage 2. The system fails when the net count of valid shocks reaches another predetermined higher threshold $m_2$ ($> m_1$).

We assume that the inter-arrival times between successive VS and those between PI are independent and follow *arbitrary* distributions $F$ and $G$, respectively. We compute the distributions of the sojourn time in Stage 1 and the failure time of the system using two approaches. We calculate the percentage improvement in the system lifetime after subdividing Stage 1. Finally, we make optimal choices which minimize the expected maintenance cost per unit time for two maintenance policies.


*Key words*: Reliability; Counting process; Weighted convolution; Mean time to failure; Replacement cost

*AMS Subject Classifications*: 90B25, 62N05, 60K10

## 1. Introduction

In machine maintenance and reliability engineering, it is often necessary to study the impacts of external shocks. Other than degradation due to aging, system lifetime is affected by the accumulated damage due to shocks. Because a system failure causes a severe loss, it is preferable to replace a system before it fails, but only after utilizing its potential life to the extent possible. Therefore, we seek optimal replacement policies before the system fails.

Shock models have been studied extensively in the past four decades. They can be

Correponding Author: Jyotirmoy Sarkar
Email: jsarkar@iupui.edu

classified according to types of shocks, arrival processes of shocks, types of systems, nature of system degradation and so on. A comprehensive literature review is given in Chatterjee and Sarkar (2021); it will not be repeated here, except to draw attention to papers dealing with self-healing such as Dong *et al.* (2020) and Shen *et al.* (2018), together with applications of self-healing mentioned in Lafont *et al.* (2012) (Solid-state lighting devices), Keedy and Feng (2013) (bio-medical application), and Bhuyan and Dewanji (2017) (secondary email account system and battery life of cell phones). Let us mention in some detail only one paper Zhao *et al.* (2018) , which is instrumental in developing our model, though we modify some of the assumptions therein to make our model more realistic.

In Zhao *et al.* (2018), the authors studied a two-stage shock model with a self healing mechanism. A shock that results in a certain degree of damage is called valid; otherwise, it is deemed invalid. A $\delta$-invalid shock is one whose time lag with the preceding shock exceeds a given time-threshold $\delta$. A change point splits the failure process into two stages: In Stage 1, $\delta$-invalid shocks trigger self healing by essentially reducing by one the number of valid shocks. When the cumulative number of net valid shocks exceeds a critical level $d$, the system enters Stage 2, where it loses its healing ability. The cumulative number of valid shocks increases by one whenever a valid shock arrives; but the self-healing behaviour is triggered only when there is a running trail of $k$ **consecutive $\delta$-invalid** shocks, reducing the cumulative number of damage by one. A system fails when the cumulative number of valid shocks reaches a threshold $n$ ($n > d$). Further, the authors considered three preventive maintenance policies and the optimal replacement strategies for each policy after considering the associated costs. In their illustrations, inter-arrival times between successive valid shocks and $\delta$-invalid shocks are exponentially distributed.

In Chatterjee and Sarkar (2021), two types of impacts — valid shocks (VS) and positive interventions (PI) — are considered, with their inter-arrival times having *arbitrary* distributions, and the system lifetime is split into two stages — Stage 1 where it can heal, and Stage 2 where it can not heal. However, healing occurs when the **cumulative** effect of $k$ PIs (not necessarily consecutive) nullify one VS. Furthermore, the PIs need not be $\delta$-invalid. This continues until the system reaches a "change point" beyond which it can no longer heal.

The main focus of the current work is to extend the two-stage model by splitting Stage 1 further into two parts. In the earlier part of Stage 1, called Stage 1A, $k_A$ PIs nullify the damaging effect of one VS. In the later part of Stage 1, called Stage 1B, $k_B$ ($> k_A$) PIs can heal one VS. The system is in Stage 1A until the net VS reaches a threshold $m_A$; thereafter, it enters Stage 1B. Next, the system reaches the "change point" and enters Stage 2 when the net VS reaches $m_1$. Therefore, $m_1 - m_A = m_B$ is the net number of VS allowed in Stage 1B.

Previous research considered either healing or degradation, or they have assumed that the shocks/impacts have inter-arrival times exponentially distributed. Although some works mention non-exponential inter-arrival times, they illustrate only exponential examples. As in Chatterjee and Sarkar (2021), here we illustrate with several non-exponential inter-arrival time distributions. As long as we can count the number of VS and PI, we can figure out the distributions of duration of Stage 1 and system lifetime.

Section 2 describes the evolution of the system under shocks and healing; Section 3

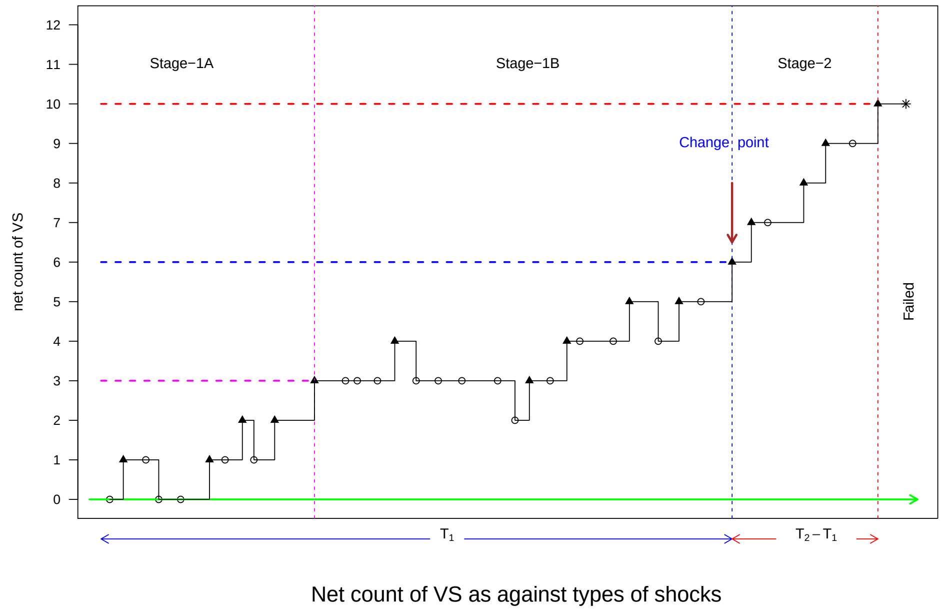


**Figure 1: Stages of the system based on net count of VS**

illustrates two approaches to compute the distributions of Stage 1 duration and lifetime; Section 4 compares Stage 1 duration and system lifetime between divided Versus undivided Stage 1; Section 5 obtains optimal decisions for two maintenance policies; and finally Section 6 summarizes the main findings of this research.

## 2. The System Set-up

External impacts to the system are of two types: **Valid Shocks (VS)** that cause damage to the system and **Positive Interventions (PI)** that do not have any damaging effect; on the contrary, the accumulation of a certain (predetermined) number of PIs nullify the effect of one VS. This behaviour is what we call **healing**, which means the net number of VS (VS arrived minus VS nullified by PIs) reduces by one. For simplicity, we assume each VS causes an equal amount of damage. Hence, leaving for future the study of magnitude of damages, here we focus on counting the net number of VS to the system.

The system lifetime is divided into two stages depending on the net VS it receives. In Stage 1, the system has healing ability as described above and the system remains in this stage until the net VS reaches a certain predetermined threshold $m_1$. Thereafter, the system moves to Stage 2 where it can no longer heal; that is, new PIs do not reduce the net VS anymore. The system fails when net VS reaches another higher threshold $m_2$. Furthermore, Stage 1 is subdivided into two parts: Stage 1A requires fewer and Stage 1B requires larger number of PI's to nullify one VS.

In Figure 1, the arrival processes of VS (denoted by ▲) and PI (denoted by ○) illustrate the net count of VS, and hence the stages. Here, $k_A = 2, k_B = 4, m_A = 3, m_1 = 6$, and

$m_2 = 10$. Do not start counting until the first VS arrives. Stop counting PI if the net number of VS, drops to 0. Resume counting once the next VS arrives. The **change point** $T_1$ defines the transition from Stage 1 (Stages 1A and 1B combined) to Stage 2.

The inter-arrival times of VS, denoted by $X_1, X_2, X_3, \ldots$ are independently and identically distributed (IID) with an *arbitrary* cumulative distribution function (CDF) $F$. Likewise, $Y_1, Y_2, Y_3, \ldots,$ the inter-arrival times of PIs are IID with another *arbitrary* CDF $G$. Arrival processes of PIs and VS are stochastically independent. Let the duration of system in Stage 1 be denoted as $T_1$ and the system lifetime be denoted as $T_2$. The total number of VS in Stage 1 be $N_1$ and that until failure be $N_2$. Note that $m_2 - m_1 = N_2 - N_1$ since in Stage 2 there is no healing. Let $r$ denote the number of PIs rendered unused towards healing in Stage 1. Further, let $D_1$ and $D_2$ denote the total number of impacts (VS+PI) in Stage 1 and until failure, respectively.

## 3. Distributions of $T_1$ and $T_2$

We describe the underlying stochastic process in terms of two approaches: a counting process approach and a convolution process approach.

### 3.1. The counting process

Given the constant integers $k_A$, $k_B$, $m_A$, $m_1$ (hence, $m_B = m_1 - m_A$) and $m_2$, we describe a simulation of the system status as follows.

Generate a sequence of inter-arrival times of VS $X_1, X_2, X_3, \ldots$ from $F$, and another sequence of inter-arrival times of PIs $Y_1, Y_2, Y_3, \ldots$ from $G$. Take the cumulative sums of the two sequences, to obtain the arrival times. Sort these arrival times of the impacts (VS and PI), and associate with each arrival time an indicator **1** to denote VS and **0** to denote PI. Begin counting as soon as the first VS arrives. Ignore all PIs (0) before this moment.

Stage 1A: Count the VS. Arrival of $k_A$ PIs nullify one VS. Compute the net VS as the VS arrived minus VS nullified. If the net VS ever drops to 0, stop counting; resume counting when again another VS arrives. When net VS reaches $m_A$, the system enters Stage 1B. We keep record of the total number of VS that arrived in Stage 1A, namely $N_A$.

Stage 1B: In this later part of Stage 1, arrival of a VS increases its count by one, but now to nullify one VS we need $k_B$ PIs ($k_B > k_A$). Again, we stop counting if the net count ever drops down to 0; and resume counting when a new VS arrives. When the net VS reaches $m_1$, the system enters Stage 2. Let $N_B$ denote the total number of VS that arrive Stage 1B, and let $r$ denote the number of PIs that have arrived in Stage 1, including those that arrived before the very first VS, those which were used for nullifying VS, those which arrived after the net count of VS dropped down to 0 and those that were unused towards healing. Similarly let $s$ denote the total number of PI that arrived in Stage 2.

Stage 2: In this stage the system does not heal. The VS keeps accumulating one by one without being nullified since the PIs have no effect. The system fails when the net VS reaches a threshold $m_2$.

Thus, in one iteration of the simulation, we obtain as outputs the following quantities:

$N_1 = N_A + N_B$, $N_2$, $T_1$, $T_2$, $r$; $D_1 = N_1 + r$ which is the total number of impacts in Stage 1 and $D_2 = N_2 + r + s$ which is total number of impacts in Stage 1 and Stage 2 combined. Next, we repeat the above steps for a total of $10^4$ iterations.

We will approximate the probability mass functions (PMF) of $N_1$ and $N_2$ from the relative frequencies observed in the simulation. Also, based on the simulation, we will directly approximate the probability density function (PDF) of $T_1$ and $T_2$. Alternatively, we will reconstruct these PDFs using the PMFs of $N_1$ and $N_2$, respectively, through a convolution process as explained below.

### 3.2. The convolution process

The underlying stochastic process is described below:

In Stage 1A, $(N_A - m_A)$ VS have been nullified by the arrival of $(N_A - m_A) * k_A$ PIs. Similarly in Stage 1B, $(N_B - m_B)$ VS have been nullified by the arrival of $(N_B - m_B) * k_B$ PI and there may have arrived $h$ more PIs, where $0 \leq h \leq k_B - 1$, which are insufficient to nullify another VS. Therefore, in total, the arrival of $Q = (N_A - m_A) * k_A + (N_B - m_B) * k_B$ PIs has contributed towards nullifying $(N_1 - m_1)$ PIs.

Let us denote $S_j = \sum_{i=1}^{j} X_i$ as arrival time of the $j$-th VS, and $U_j = \sum_{i=1}^{j} Y_i$ as arrival time of the $j$-th PI. We describe how Stage 1 duration $T_1$ depends on the number of VS $N_1$.

(1) The system receives $N_1$ VS in Stage 1. The arrival time of the $(N_1 - 1)$-st VS is $S_{N_1-1}$ and that of the $N_1$-th VS is $S_{N_1}$.

(2) Let $U_Q$ be the arrival time of a PI which causes the $(N_1 - m_1)$-th nullification of a VS, and let $U_{Q+h}$ be the arrival time of the $(Q + h)$-th PI, which do not nullify any VS (where $h = 1, 2, \ldots, k_2 - 1$), since the count of unused PI has not reached $k_B$ yet.

(3) Before the $(N_1 - 1)$-st VS arrives, the $Q$-th PI has already arrived. Hence, $U_Q < S_{N_1-1}$. Thereafter, until the $N_1$-th VS arrives, fewer than $k_B$ PI have arrived in Stage 1B. Therefore, the arrival times satisfy the inequality

$$U_Q < S_{N_1-1} < U_{Q+h} < S_{N_1} < U_{Q+k_2} \tag{1}$$

and the sojourn time in Stage 1 is

$$T_1 = S_{N_1} \tag{2}$$

Note that $S_j$ has a CDF given by $F * F * \cdots * F$, the $j$-fold convolution of $F$. Moreover, since $N_1$ is a random stopping time (that determines the end of Stage 1), by Wald's first identity, we have $E(T_1) = E(N_1) \times E(X)$. However, $N_1$ and $T_1$ are not independent. Therefore, using a second-order approximation (by matching the mean and the mean squared deviation from the mean), we model

$$T_1 = S_j + \lambda \left( j - E[N_1] \right) E[X] \text{ with probability } P(N_1 = j) \tag{3}$$

for $j = m_1, m_1 + 1, \ldots$; where $\lambda \in [0, 1]$ depends on $F$ and $G$. That is, the distribution of $T_1$ is modeled as a weighted average of adjusted $j$-fold convolutions of $F$, where the adjustment equals a suitable fraction of the departure of $N_1$ from its expectation *times* the expected inter-arrival time between shocks, with weights given by the probability masses $P(N_1 = j)$ for $j = m_1, m_1 + 1, \ldots$.

The above explanations justify the following theorem:

**Theorem 1:** The distribution of Stage 1 duration is a weighted average of $j$-fold convolutions of $F$ shifted by $\lambda\,(j - E[N_1])\,E[X]$, where $\lambda \in [0, 1]$ is described below, with weights given by $P\{N_1 = j\}$, the probability that $N_1$ VS arrive in Stage 1, for $j = m_1, m_1 + 1, \ldots$.

*Description of $\lambda$:* For several combinations of inter-arrival time distributions $F$ and $G$, the fraction $\lambda$ is numerically obtained via a grid search (with increment 0.01) to match the standard deviations of the distribution of $T_1$ obtained from the point process and the convolution process.

The system lifetime equals the duration of Stage 1 *plus* $(m_2 - m_1)$ additional inter-arrival times of VS (which is the duration of Stage 2), since the system can no longer heal in Stage 2. Hence, the system lifetime is

$$T_2 = T_1 + \sum_{i=N_1+1}^{N_1+m_2-m_1} X_i = S_{N_2}$$

where $N_2 = N_1 + m_2 - m_1$. Using Theorem 1, we can describe

$$T_2 = S_{j+m_2-m_1} + \lambda\,(j - E[N_1])\,E[X] \text{ with probability } P\{N_1 = j\} \tag{4}$$

for $j = m_1, m_1 + 1, \ldots$.

**Corollary 1:** The distribution of time to failure $T_2$ is a weighted average of $(j + m_2 - m_1)$-fold convolution of $F$ shifted by $\lambda\,(j - E[N_1])\,E[X]$, with weights given by $P\{N_1 = j\}$, for $j = m_1, m_1 + 1, \ldots$.

It is noteworthy that the above theorem and corollary are exactly the same as in Chatterjee and Sarkar (2021). The intuition behind this agreement is that both the theorem and the corollary involve the PMF of $N_1 = N_A + N_B$, the total number of VS in stages 1A and 1B combined. Exactly how Stage 1 is subdivided (based on the requirements of healing) is irrelevant to describe the Stage 1 duration or the lifetime.

### 3.3. Computation and comparison

We shall consider different inter-arrival time distributions for $X$ and $Y$ satisfying $E(X) = 1$ and $E(Y) = 2/3$. The distribution of sojourn time in Stage 1 is found directly by repeating the point process $10^4$ times. In Chatterjee and Sarkar (2021), we had considered $k = 3$ and $m_1 = 10$ for illustration. For comparability, here we choose $k_A = 2, k_B = 4$ to keep the overall average number of PIs required to nullify one VS roughly the same and we choose $m_A = 5, m_B = 5$, so that $m_1 = 10$.

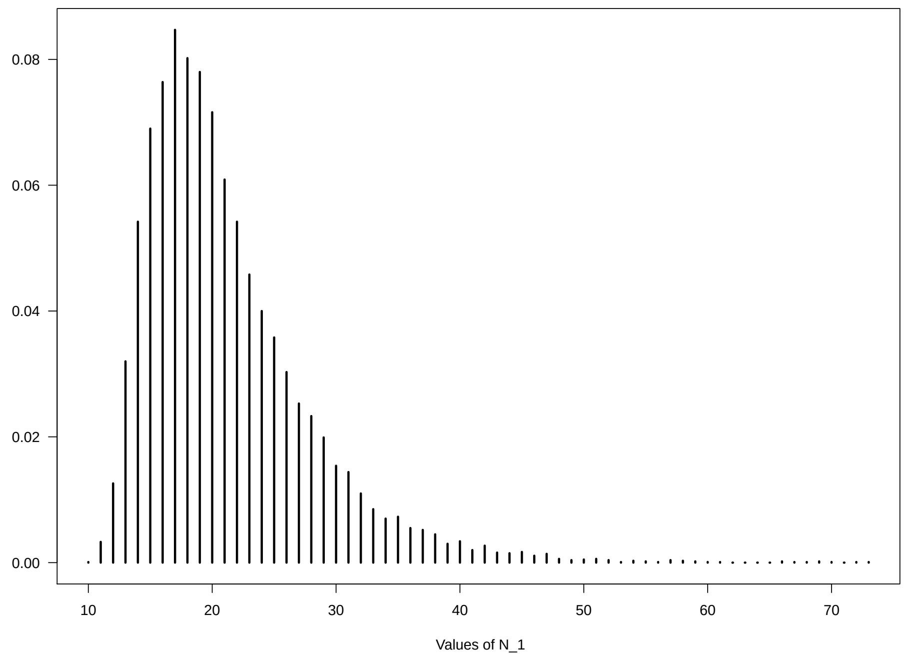


**Figure 2: Probability Distribution of $\mathbf{N_1}$ is unimodal with mode 17, $\mathbf{E(N_1) \approx 21}$, $\mathbf{SD(N_1) = 6.52}$, $\mathbf{Q_1 = 17}$, $\mathbf{Q_2 = 20}$, $\mathbf{Q_3 = 24}$, 99-th percentile $\mathbf{(N_1)_{0.99} = 36, P(N_1 > 40) = 0.0057}$.**

We emphasize that while similar results hold for all combinations of inter-arrival times, to save space, we will show detailed results for one particular combination of inter-arrival times: $F \equiv$ Weibull (shape $= 2$, scale $= 2/\sqrt{\pi}$) and $G \equiv$ Gamma (shape $= 2$, scale $= 1/3$), such that $E(X) = 1$ and $E(Y) = 2/3$. Figure 3 shows the simulated PMF of $N_1$.

Figure 3 shows the PDFs of $T_1$ and $T_2$ obtained both directly from point process and from adjusted convolution. For $T_1$, the mean Kullback-Leibler divergence of the adjusted convolution PDF from the point process PDF measures 0.001466, which is very small (with simulated p-value 0.997) and supports Theorem 1. Similarly, for $T_2$, the mean Kullback-Leibler divergence of 0.00102 (with simulated p-value 0.999) supports Corollary 1. The discrepancies between the two PDFs is at most 0.00035, establishing that the PDFs arising from the point process and the adjusted convolution process are the same,

Let us now consider all combinations of $F$ and $G$ simultaneously. In the Table 1, we show the mean and the standard deviations of $T_1$ obtained from the two processes for various choices of $F$ and $G$. Similarly, in Table 2, we show the mean and the standard deviations of $T_2$. We show the corresponding $\lambda$'s for each combination of $F$ and $G$ in Table 3.

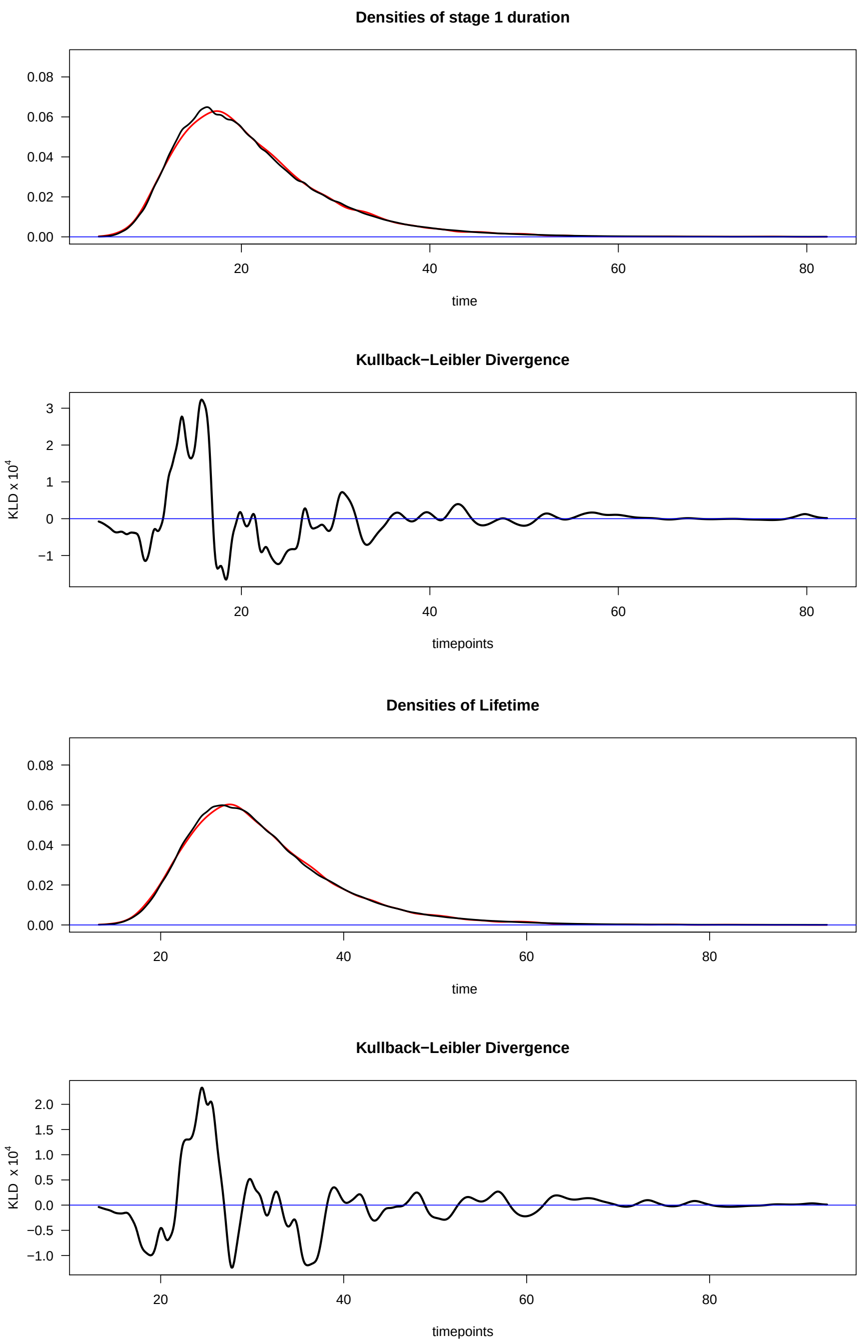


**Figure 3: Densities of $T_1$ (upper) and $T_2$ (lower) estimated from a point process (red) and an adjusted convolution process (black), with their difference being within $3.5 \times 10^{-4}$ of 0.**

**Table 1: For various inter-arrival time distributions satisfying $\mathbf{E(X) = 1}$ and $\mathbf{E(Y) = 2/3}$ the top entries give mean (standard deviation) of Stage 1 duration $\mathbf{T_1}$ according to a point process, and the bottom entries (in *italics*) show the same quantities according to an adjusted convolution process.**

| VS \ PI | Weibull $(2, \frac{4}{3\sqrt{\pi}})$ $SD \approx 0.12$ | Gamma $(2, \frac{1}{3})$ $SD \approx 0.22$ | Inv-Gauss $(2/3)$ $SD \approx 0.29$ | Exponential $(3/2)$ $SD \approx 0.40$ |
|---|---|---|---|---|
| Weibull $(2, \frac{2}{\sqrt{\pi}})$ | 21.51 (8.12) | 21.17 (8.35) | 21.01 (8.50) | 20.34 (8.39) |
| $SD \approx 0.27$ | *21.50 (8.15)* | *21.15 (8.38)* | *21.01 (8.53)* | *20.31 (8.39)* |
| Gamma $(2, \frac{1}{2})$ | 20.77 (9.13) | 20.42 (9.07) | 20.43 (9.21) | 19.82 (9.05) |
| $SD \approx 0.50$ | *20.76 (9.12)* | *20.44 (9.07)* | *20.43 (9.20)* | *19.81 (9.01)* |
| Inv-Gauss (1) | 19.53 (10.16) | 19.26 (10.01) | 19.18 (10.11) | 18.88 (9.97) |
| $SD \approx 1.00$ | *19.61 (10.16)* | *(19.36 (10.01)* | *19.27 (10.07)* | *18.98 (9.97)* |
| Exponential (1) | 19.59 (10.36) | 19.41 (10.36) | 19.46 (10.59) | 18.91 (10.04) |
| $SD \approx 1.00$ | *19.59 (10.37)* | *19.37 (10.39)* | *19.43 (10.58)* | *18.89 (10.06)* |

## 4. Comparison with Undivided Stage 1

In Table 3, we compare the means of the Stage 1 duration, showing the percentage change, between the divided Stage 1 studied here and the undivided Stage 1 studied in Chatterjee and Sarkar (2021). We also report the $\lambda$'s obtained in the current research and compare them to the $\lambda$'s reported in Chatterjee and Sarkar (2021).

From Table 3, we identify a trend in the values of $\lambda$ as we scan through the rows and the columns. For a particular choice of $F$ in a row, as we look from left to right across the columns, we see that $\lambda$ decreases. A closer look at the corresponding standard deviations reveals that $\lambda$ decreases as the standard deviation of $G$ increases. Similarly, for a fixed choice of $G$ in a particular column, as we go from top to bottom down the rows, we see that $\lambda$ increases as the standard deviation of $F$ increases. This led us to believe that $\lambda$ is a function of the ratio of the standard deviations of $F$ and $G$. When we plotted $\lambda$ against $\sigma_F/\sigma_G$, we noticed a non-linear relationship. Thereafter, we fitted a linear regression of $\lambda$ on $\log(\sigma_F/\sigma_G)$ with slope = 0.11833, intercept = 0.20644 and adjusted coefficient of determination of 0.832.

Let us look at the change in $\lambda$ before and after the subdivision of Stage 1. When the SD of $F$ is $\approx 0.25$ or $\approx 0.50$, the $\lambda$'s for the divided Stage 1 is about one-half to three-fifths of the $\lambda$'s from the undivided Stage 1; but when the SD of $F$ is $\approx 1$, the $\lambda$'s for the divided Stage 1 is about two-thirds of the $\lambda$'s from the undivided Stage 1.

In Table 4, we compare the means of the lifetimes, showing the percentage change, between the divided Stage 1 studied here and the undivided Stage 1 studied in Chatterjee and Sarkar (2021).

**Table 2: For various inter-arrival time distributions satisfying $\mathbf{E(X) = 1}$ and $\mathbf{E(Y) = 2/3}$ the top entries give mean (standard deviation) of system lifetime $\mathbf{T_2}$ according to a point process, and the bottom entries (in *italics*) show the same quantities according to an adjusted convolution process.**

| VS \ PI | Weibull $(2, \frac{4}{3\sqrt{\pi}})$ $SD \approx 0.12$ | Gamma $(2, \frac{1}{3})$ $SD \approx 0.22$ | Inv-Gauss (2/3) $SD \approx 0.29$ | Exponential (3/2) $SD \approx 0.40$ |
|---|---|---|---|---|
| Weibull $(2, \frac{2}{\sqrt{\pi}})$ | 31.48 (8.27) | 31.14 (8.51) | 31.01 (8.65) | 30.32 (8.58) |
| $SD \approx 0.27$ | *31.50 (8.31)* | *31.15 (8.55)* | *31.03 (8.69)* | *30.33 (8.55)* |
| Gamma $(2, \frac{1}{2})$ | 30.77 (9.44) | 30.44 (9.34) | 30.42 (9.48) | 29.82 (9.34) |
| $SD \approx 0.50$ | *30.76 (9.39)* | *30.45 (9.33)* | *30.43 (9.47)* | *29.80 (9.28)* |
| Inv-Gauss (1) | 29.45 (10.69) | 29.23 (10.53) | 29.17 (10.66) | 28.83 (10.54) |
| $SD \approx 1.00$ | *29.55 (10.67)* | *29.32 (10.51)* | *29.23 (10.57)* | *28.91 (10.49)* |
| Exponential (1) | 29.62 (10.80) | 29.42 (10.82) | 29.51 (11.07) | 28.95 (10.53) |
| $SD \approx 1.00$ | *29.57 (10.89)* | *29.34 (10.88)* | *29.43 (11.08)* | *28.90 (10.58)* |

## 5. Preventive Maintenance Policies

System failure being disruptive to the production process and too expensive to recover from, oftentimes a maintenance engineer must intervene to replace a functioning unit. Clearly there is a tension between utilizing the remaining lifetime of the system and the prevention of failure. We consider here two types of preventive maintenance policies. Let $c_{p_A}$ be the cost of replacement in Stage 1A, $c_{p_B}$ in Stage 1B, $c_{p_2}$ in Stage 2, and $c_f$ after failure. Furthermore, we assume that the costs of replacement is the same throughout Stage 1, because the healing rate ought not affect the cost of replacement. We consider $c_{p_A} = c_{p_B} < c_{p_2} << c_f$ with justification as follows: In Stage 1, the system is young, and so an early replacement will incur a loss; if we replace in early part of Stage 2, we are not utilizing the system lifetime sufficiently, but the system has already aged, and so maintenance/repair at this stage will cost more. Therefore, we find it logical to consider that replacement cost in Stage 2 is higher than that in Stage 1. Finally, a system failure is highly expensive. Furthermore, there is an initial cost $c_0$ of setting up a new system. For illustration, we choose $c_0 = 100, c_{p_A} = 10, c_{p_B} = 10, c_{p_2} = 15, c_f = 200$. In Figures 4 and 5, we illustrate decision making when $F \equiv$ Weibull $(2, 2/\sqrt{\pi})$ and $G \equiv$ Gamma (2, 1/3), such that $E(X) = 1$ and $E(Y) = 2/3$.

### 5.1. Maintenance Policy 1

Suppose that a monitoring equipment can detect the arrival of an impetus, but cannot distinguish between a VS and a PI, nor can it identify whether the system is in Stage 1 or Stage 2. The system will be replaced when it has failed or has experienced a specified number of impetus $N$ (the sum of VS and PI).

Within one cycle (between two successive replacements of the system), the total cost of replacement $C_1$ under Policy 1 is a random variable taking three possible values:

**Table 3: For various inter-arrival time distributions satisfying $\mathbf{E(X) = 1}$ and $\mathbf{E(Y) = 2/3}$, the top entries give the mean duration of $\mathbf{T_1}$ for divided Stage 1 (undivided Stage 1), the middle row gives the % increase in $\mathbf{T_1}$ after subdividing Stage 1, and the third row gives the multiplier $\lambda$ of adjusted convolution for the divided Stage 1 (undivided Stage 1).**

| VS \ PI | Weibull $(2, \frac{4}{3\sqrt{\pi}})$ $SD \approx 0.12$ | Gamma $(2, \frac{1}{3})$ $SD \approx 0.22$ | Inv-Gauss $(2/3)$ $SD \approx 0.29$ | Exponential $(3/2)$ $SD \approx 0.40$ |
|---|---|---|---|---|
| Weibull $(2, \frac{2}{\sqrt{\pi}})$ $SD \approx 0.27$ | 21.51 (17.96) $\approx 19.76\%$ $\lambda = 0.25(0.50)$ | 21.17 (17.97) $\approx 17.8\%$ $\lambda = 0.21(0.40)$ | 21.01 (17.98) $\approx 16.85\%$ $\lambda = 0.18(0.35)$ | 20.34 (17.93) $\approx 13.44\%$ $\lambda = 0.16(0.29)$ |
| Gamma $(2, \frac{1}{2})$ $SD \approx 0.50$ | 20.77 (17.92) $\approx 15.90\%$ $\lambda = 0.31(0.59)$ | 20.42 (17.91) $\approx 14.00\%$ $\lambda = 0.27(0.50)$ | 20.43 (17.93) $\approx 13.94\%$ $\lambda = 0.25(0.46)$ | 19.82 (17.85) $\approx 11.03\%$ $\lambda = 0.22(0.38)$ |
| Inv-Gauss (1) $SD \approx 1.00$ | 19.53 (17.63) $\approx 10.78\%$ $\lambda = 0.43(0.68)$ | 19.26 (17.61) $\approx 9.37\%$ $\lambda = 0.40(0.61)$ | 19.18 (17.67) $\approx 8.55\%$ $\lambda = 0.37(0.58)$ | 18.88 (17.61) $\approx 7.21\%$ $\lambda = 0.36(0.52)$ |
| Exponential (1) $SD \approx 1.00$ | 19.59 (17.85) $\approx 9.75\%$ $\lambda = 0.45(0.69)$ | 19.41 (17.87) $\approx 8.62\%$ $\lambda = 0.43(0.63)$ | 19.46 (17.92) $\approx 8.59\%$ $\lambda = 0.39(0.60)$ | 18.91 (17.82) $\approx 6.12\%$ $\lambda = 0.37(0.54)$ |

**Table 4: For various inter-arrival time distributions satisfying $\mathbf{E(X) = 1}$ and $\mathbf{E(Y) = 2/3}$, the top row gives the mean duration of $\mathbf{T_2}$ for divided Stage 1 (undivided Stage 1), the bottom row gives approximate % increase in mean $\mathbf{T_2}$ after subdividing Stage 1.**

| VS \ PI | Weibull $(2, \frac{4}{3\sqrt{\pi}})$ $SD \approx 0.12$ | Gamma $(2, \frac{1}{3})$ $SD \approx 0.22$ | Inv-Gauss $(2/3)$ $SD \approx 0.29$ | Exponential $(3/2)$ $SD \approx 0.40$ |
|---|---|---|---|---|
| Weibull $(2, \frac{2}{\sqrt{\pi}})$ $SD \approx 0.27$ | 31.48 (27.94) $\approx 12.67\%$ | 31.14 (27.96) $\approx 11.37\%$ | 31.01 (27.96) $\approx 10.91\%$ | 30.02 (27.92) $\approx 7.52\%$ |
| Gamma $(2, \frac{1}{2})$ $SD \approx 0.50$ | 30.77 (27.91) $\approx 10.25\%$ | 30.44 (27.92) $\approx 9.03\%$ | 30.42 (27.93) $\approx 8.92\%$ | 29.82 (27.86) $\approx 7.03\%$ |
| Inv-Gauss (1) $SD \approx 1.00$ | 29.45 (27.57) $\approx 6.82\%$ | 29.23 (27.58) $\approx 5.98\%$ | 29.17 (27.60) $\approx 5.69\%$ | 29.83 (27.54) $\approx 8.32\%$ |
| Exponential (1) $SD \approx 1.00$ | 29.62 (27.86) $\approx 6.32\%$ | 29.42 (27.90) $\approx 5.45\%$ | 29.51 (27.93) $\approx 5.66\%$ | 28.95 (27.82) $\approx 4.06\%$ |

(1) $c_{p_1}$, if $N$ impetus arrive while the system is still in Stage 1, with an associated probability of $P(D_1 > N)$.

(2) $c_{p_2}$, if the system has already moved from Stage 1 to Stage 2 and the $N$ impetus have arrived before system failure, with an associated probability of $P(D_1 \le N < D_2)$.

(3) $c_f$, if the system has already failed before the arrival of $N$ impetus, with an associated probability of $P(D_2 \le N)$.

Hence, the expected cost (C) under Policy 1, is given by

$$E(C|\text{Policy 1}) = c_{p_1}\, P(D_1 > N) + c_{p_2}\, P(D_1 \le N < D_2) + c_f\, P(D_2 \le N) \tag{5}$$

and, writing $W_j$ as the arrival time of the $j$-th impact (either VS or PI), the expected cycle time (CT) is given by

$$\begin{aligned} E(CT|\text{Policy 1}) = E[\min\{W_N, W_{D_2}\}] &= E[W_N|D_1 > N]\ P(D_1 > N) \\ &+ E[W_N|D_1 \le N < D_2]\ P(D_1 \le N < D_2) \\ &+ E[W_{D_2}|D_2 \le N]\ P(D_2 \le N). \end{aligned} \tag{6}$$

Therefore, the expected cost per unit time is the ratio

$$E(C|\text{Policy 1})/E(CT|\text{Policy 1}) \tag{7}$$

which we must minimize by choosing $N$.

For the example considered, Figure 4 shows that the expected cost per unit time is minimized when we choose $N = 56$. Moreover, note that for any other choice of $N$ in the vicinity of the optimal value 56, say between 50 and 60, the expected cost per unit time increases only slightly (no more than 3%). Such a robustness result allows us to rely on the optimal value even when the inter-arrival time distributions deviate slightly from the stated ones. Table 5 documents the optimal choices for other combinations of inter-arrival times.

### 5.2. Maintenance Policy 2

Suppose that the monitoring equipment can identify the stages of the system. If the system is in Stage 1, we do not replace it at all. After the system enters Stage 2, if the system is still functioning for an additional $t$ units of time, we replace it immediately at epoch $T_1+t$; otherwise, we replace the system immediately on failure during $[T_1, T_1+t)$. Our objective is to determine an optimum additional time $t$ in Stage 2. To do so, we minimize the expected cost per unit time, where the expected cost (C) is given by

$$E(C|\text{Policy 2}) = c_{p_2} P(T_2 > T_1 + t) + c_f P(T_2 \le T_1 + t) \tag{8}$$

and the expected length of the cycle time (CT) is

$$E(CT|\text{Policy 2}) = E(min(T_2, T_1 + t)) = E(T_1) + E(min(T_2 - T_1, t)) \tag{9}$$

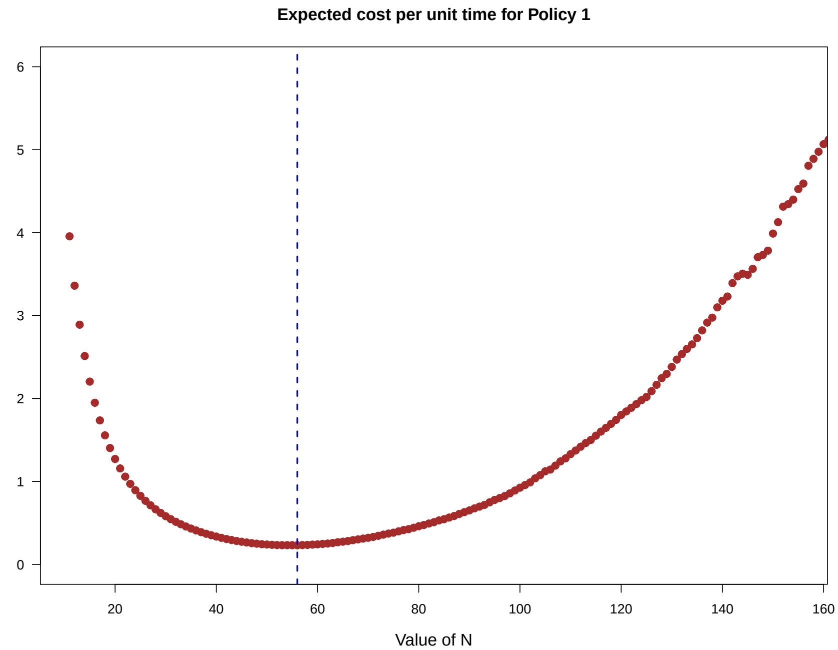


**Figure 4: Under Policy 1, when $\mathbf{F \equiv}$ Weibull (2, $\mathbf{2/\sqrt{\pi}}$) and $\mathbf{G \equiv}$ Gamma (2, 1/3), and cost parameters are $\mathbf{c_0 = 100}$, $\mathbf{c_{p_{1a}} = 10}$,$\mathbf{c_{p_{1b}} = 10}$, $\mathbf{c_{p_2} = 15}$, $\mathbf{c_f = 200}$, the maintenance cost per unit time is minimized when we choose $\mathbf{N = 56}$.**

We wish to minimize the expected cost per unit time

$$E(C|\text{Policy 2})/E(CT|\text{Policy 2}) \tag{10}$$

by choosing $t$. Under Policy 2, the assumed cost parameters, and $F \equiv$ Weibull (2, $2/\sqrt{\pi}$) and $G \equiv$ Gamma (2, 1/3), Figure 5 shows that the expected cost per unit time is minimized at $t = 6.6$. In fact, we identified this optimal $t$ value via a grid search between the first and the 99-th percentiles of system lifetime with an increment of 0.05. This choice suffices because any other choice of $t$ in the interval $[6, 7]$ increases the cost per unit time only marginally.

Table 5 gives the summary of the optimal choices of $N$ and $t$ for Policy 1 and Policy 2, respectively, for different choices of $F$ and $G$ and for cost parameters $c_0 = 100, c_{p_{1a}} = 10, c_{p_{1b}} = 10, c_{p_2} = 15, c_f = 200$.

We see that for policy 1, the total number of impacts $N$ is only 0-3 impacts more than the optimal values of $N$ when Stage 1 was not divided. This close agreement should not come as a surprise because, for our choice of $(k_A, k_B)$ and $(m_A, m_B)$, the average impact throughout the entire undivided Stage 1 is comparable to that in the undivided Stage 1 case. Hence, there is only a negligible amount of change in $N$ due to subdivision. Similarly, for policy 2, there is no significant change in $t$, because the choice of $t$ depends only on the arrival rate of VS in Stage 2 and not at all on the subdivision of Stage 1 to accommodate varying rates of healing.

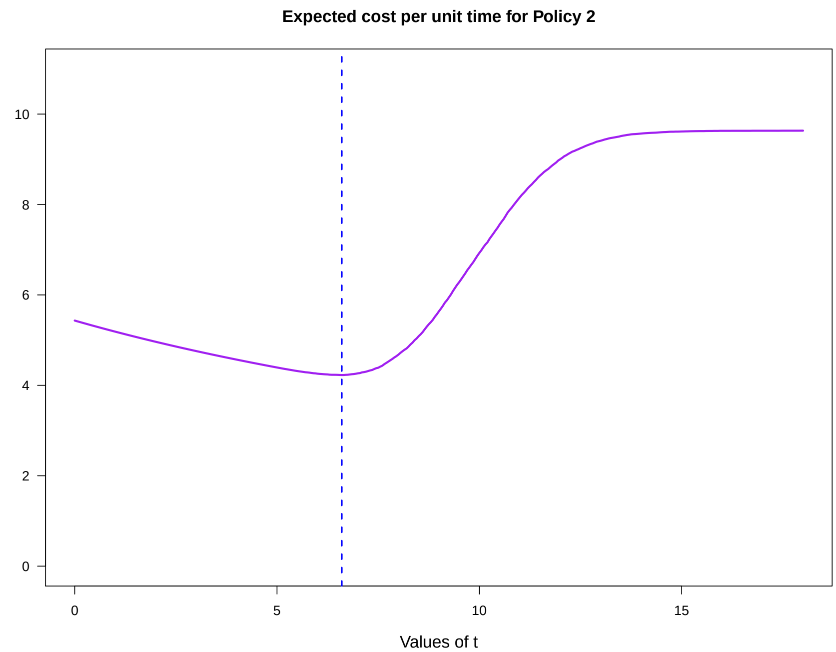


**Figure 5: Under Policy 2, when $F \equiv$ Weibull $(2, 2/\sqrt{\pi})$, $G \equiv$ Gamma $(2, 1/3)$, and cost parameters are $c_0 = 100$, $c_{p_{1a}} = 10$, $c_{p_{1b}} = 10$, $c_{p_2} = 15$, $c_f = 200$, the maintenance cost per unit time is minimized when we choose $t = 6.6$.**

**Table 5: For various inter-arrival time distributions F and G satisfying $E(X) = 1$ and $E(Y) = 2/3$, to minimize the maintenance cost per unit time, the optimal N for Policy 1 is shown in the first row, and the optimal t for Policy 2 is shown in the second row.**

| VS \ PI | Weibull $(2, \frac{4}{3\sqrt{\pi}})$ $SD \approx 0.12$ | Gamma $(2, \frac{1}{3})$ $SD \approx 0.22$ | Inv-Gauss $(2/3)$ $SD \approx 0.29$ | Exponential $(3/2)$ $SD \approx 0.40$ |
|---|---|---|---|---|
| Weibull $(2, \frac{2}{\sqrt{\pi}})$ $SD \approx 0.27$ | $N = 55$ $t = 6.45$ | $N = 56$ $t = 6.60$ | $N = 53$ $t = 6.55$ | $N = 52$ $t = 6.60$ |
| Gamma $(2, \frac{1}{2})$ $SD \approx 0.50$ | $N = 54$ $t = 5.70$ | $N = 51$ $t = 5.75$ | $N = 52$ $t = 5.80$ | $N = 50$ $t = 5.85$ |
| Inv-Gauss $(1)$ $SD \approx 1.00$ | $N = 51$ $t = 4.85$ | $N = 50$ $t = 4.65$ | $N = 49$ $t = 5.05$ | $N = 49$ $t = 4.90$ |
| Exponential $(1)$ $SD \approx 1.00$ | $N = 50$ $t = 4.75$ | $N = 50$ $t = 4.65$ | $N = 50$ $t = 4.65$ | $N = 49$ $t = 4.80$ |

## 6. Summary and Future Work

In this paper, we subdivided Stage 1 into two parts: initially the system heals at a faster rate requiring fewer PIs to nullify one VS; but once enough net VS have accumulated, more PIs are needed to nullify one VS. We derived the distributions of Stage 1 duration and the system lifetime for any *arbitrary* inter-arrival time distributions of VS and PI, in contrast to only exponential distributions commonly assumed in the literature. Given a prefixed net number of shocks that the system can withstand in various stages, we can work out the distributions of Stage 1 duration $T_1$ and lifetime $T_2$ using a point process or an adjusted convolution process with an adjustment factor $\lambda$ that is approximately a linear function of the logarithm of the ratio of standard deviations of $F$ and $G$. The theoretical investigation of $\lambda$ remains an open problem. Moreover, in this research we found that subdivision of Stage 1 leads to an increase in the Stage 1 duration, and hence the system lifetime.

In future, we like to consider varying magnitudes of VS and PI and allow natural system degradation according to some stochastic process. Our simple counting approach suffices to make optimal decisions that minimize maintenance costs per unit time for various inter-arrival distributions and various maintenance policies with different costs of replacement in different stages. This realistic set up promotes an utmost utilization of system lifetime.

## Acknowledgements

We thank our colleagues for some discussions and feedback.

## References

Bhuyan, P. and Dewanji, A. (2017). Estimation of reliability with cumulative stress and strength degradation. *Statistics*, **51(4)**, 766–781. DOI:10.1080/02331888.2016.1277224

Chatterjee, D. and Sarkar, J. (2021). Optimal replacement policies for systems under sporadic impacts that inflict damage or trigger healing. *Submitted*

Dong, W., Liu, S., Cao, Y., Javed, S. A. and Du, Y. (2020). Reliability modeling and optimal random preventive maintenance policy for parallel systems with damage self-healing. *Computers & Industrial Engineering*, **142**. DOI:10.1016/j.cie.2020.106359

Keedy, E. and Feng, Q. (2013). Reliability analysis and customized preventive maintenance policies for stents with stochastic dependent competing risk processes. *IEEE Transactions on Reliability*, **62(4)**, 887–897. DOI:10.1109/TR.2013.2285045

Lafont, U., van Zeijl, H. and van der Zwaag, S. (2012). Increasing the reliability of solid state lighting systems via self-healing approaches: A review. *Microelectronics Reliability*, **52(1)**, 71–89. DOI:10.1016/j.microrel.2011.08.013

Shen, J., Cui, L. and Yi, H. (2018). System performance of damage self-healing systems under random shocks by using discrete state method. *Computers & Industrial Engineering*, **125**, 124–134. DOI:10.1016/j.cie.2018.08.013

Zhao, X., Guo, X. and Wang, X. (2018). Reliability and maintenance policies for a two-stage shock model with self-healing mechanism. *Reliability Engineering & System Safety*, **172**, 185–194. DOI:10.1016/j.ress.2017.12.013