ON SOME OPTIMAL SWITCHING STRATEGIES

M.S. FINKELSTEIN
Department of Mathematical Statistics
University of the Free State
PO Box 339, 9300 Bloemfontein, Republic of South Africa

ABSTRACT

Systems of components with increasing failure rates are considered. Switching from an operating component on standby can be performed at each instant of time. Optimal switching strategies, maximizing the time to the first failure of a component and to the total failure of a system are investigated. A new type of a strategy: the limit strategy is introduced. It is proved that this strategy is optimal, when there is no additional information on the state of the system. Some simple examples are considered.

Keywords: Increasing failure rate, Optimal switching strategy, Information based repair.

1. INTRODUCTION: TWO COMPONENTS

Consider a system of two identical components. Assume that the distribution function (DF) of a component time to failure $F(t)$ is absolutely continuous and the corresponding failure rate $\lambda(t)$ is increasing. Thus, $F(t) \in$ IFR (increasing failure rate) [1]. The system is non-repairable. The first component starts operating at $t = 0$, while the second one is switched off ("cold" standby). After a failure of the first component, the second starts operating immediately. The foregoing defines a usual strategy of switching upon failure.

Assume now that switching from the operating component to the standby can be performed instantaneously at any moment of time. The former operating component then starts to be the standby and vice versa, thus defining a preventive switching strategy. As will be shown, this strategy can be very effective in various situations when we want to decrease the probability of a "forced" switching upon a component's failure. A practical example is of a system with an "unreliable" switching device. If, in this case, a failure of a component occurs and the

switching device is in the failed state, then a failure of the whole system occurs. Preventive switching can decrease the probability of this event.

Therefore it is worth finding a switching strategy that, under given assumptions, maximizes the probability of the system's performance without any failure of components in a given time interval. Of course, a strategy of this kind cannot change the probability of system's failure (when both components are failed), but extending the operation period without failures of components can in itself be very important in various applications.

Denote by $P_1(S,t)$ the probability of system's performance without any failure of components in $(0,t]$ under an arbitrary switching strategy S. Let $P_1(a,t)$ denote this probability with only one switch taking place at $t' = a$. Then it is easy to prove that

$$\max{}_S (P_1(S,t)) = P_1\left(a = \frac{t}{2}, t\right). \tag{1}$$

This simply follows from the fact that, assuming only one switch at $a \in (0,t]$:

$$P_1(a,t) = \exp\left\{-\int_0^a \lambda(u)du\right\}\exp\left\{-\int_0^{t-a} \lambda(u)du\right\}.$$

In order to maximise this probability, we need

$$\min{}_a \left\{\int_0^a \lambda(u)du + \int_0^{t-a} \lambda(u)du\right\}. \tag{2}$$

After differentiating the sum of integrals in (2) and equating the result to zero, we arrive at

$$\lambda(a) = \lambda(t-a). \tag{3}$$

Equation (3) has a trivial unique solution for increasing functions, namely: $a = t/2$. Due to the additivity of the integral an additional switching cannot improve this result: the total functioning time of each component should be scheduled as $t/2$.

Denote by S_l, where "l" stands for "limit", the following strategy. Let the first component function in $(0, \Delta t]$, then the second component in $(\Delta t, 2\Delta t]$, then again the first component in $(2\Delta t, 3\Delta t]$ etc. The last step in constructing this strategy is to perform the limit transition as $\Delta t \to 0$. Denote by $\lambda_{\Delta t}(t)$, $F_{\Delta t}(t)$ and $\lambda_l(t)$, $F_l(t)$ the corresponding failure rate and the DF

of the time to the first failure of a component for the case of finite Δt and for $\Delta t \to 0$, respectively. The failure rate $\lambda_{\Delta t}(t)$ has ordinary points of discontinuity at $\Delta t, 3\Delta t, 5\Delta t,...$ .The failure rate $\lambda_l(t)$ has an infinite (countable) number of ordinary points of discontinuity and it is integrable (with respect to Lebesgue measure). Thus, $F_l(t)$ is well defined. Consider the following piecewise constant function, which constitutes an upper bound for $\lambda_{\Delta t}(t)$:

$$\hat{\lambda}_{\Delta t}(t) = \begin{cases} \lambda(\Delta t), & t \in (0, 2\Delta t] \\ \lambda(2\Delta t), & t \in (2\Delta t, 4\Delta t] \\ \lambda(3\Delta t), & t \in (4\Delta t, 6\Delta t] \\ . \qquad . \qquad . \qquad . \qquad . \end{cases} \tag{4}$$

while the lower bound is:

$$\breve{\lambda}_{\Delta t}(t) = \begin{cases} 0, & t \in (0, 2\Delta t] \\ \lambda(\Delta t), & t \in (2\Delta t, 4\Delta t] \\ \lambda(2\Delta t), & t \in (4\Delta t, 6\Delta t] \\ . \qquad . \qquad . \qquad . \qquad . \end{cases} \tag{5}$$

This means that

$$\int_0^t \breve{\lambda}_{\Delta t}(u)du < \int_0^t \lambda_{\Delta t}(u)du < \int_0^t \hat{\lambda}_{\Delta t}(u)du$$

and the same inequalities are valid for the limit case when $\Delta t \to 0$. Moreover, it can easily be seen from definitions (4) and (5) that as $\Delta t \to 0$

$$\left| \lambda_{\Delta t}(t) - \lambda\left(\frac{t}{2}\right) \right| \to 0$$

uniformly in each bounded interval. Finally

$$\lambda_l(t) = \lambda\left(\frac{t}{2}\right), \tag{6}$$

which leads to the anticipated weak convergence result:

$$\left| F_{\Delta t}(t) - F_l(t) \right| \to 0 \; as \; \Delta t \to 0 \Rightarrow F_l(t) = 1 - \exp\left\{ -\int_0^t \lambda\left(\frac{u}{2}\right)du \right\} \tag{7}$$

Actually, relation (7) defines a simple scale transformation of $F(t)$ with a scaling factor $1/2$, and the foregoing mathematical proof is a justification of this intuitively clear statement.

Hence, by means of S_l an "integrated" system of two components had been constructed. It is obvious that

$$P_1(S_l, t) = P_1\left(\frac{t}{2}, t\right), \tag{8}$$

because the total scheduled time of each component's operation for both strategies is equal to $t/2$. What is the reason for considering the limit strategy instead of the simplest with one switching? It can be easily shown that we must use S_l for maximizing the mean time to the first failure, $T_1(S)$, and this can be very important in various applications. Indeed,

$$\max{}_S \; T_1(S) = \max{}_S \; \int_0^\infty P_1(S, u) du = \int_0^\infty \exp\left\{-\int_0^t \lambda\left(\frac{u}{2}\right) du\right\} dt , \tag{9}$$

as S_l maximizes $P_1(u)$ for each $u \in (0, \infty)$. Hence, time to the first failure is stochastically larger [4,5] under the optimal strategy then under any other.

Another situation when only S_l is optimal in the above sense is when the operation of our system of two components can be terminated by some external random event. It may be some device in series with the system, for instance. It is clear that S_l maximizes the probability of system's performance without failures of components (and the time to the first failure as well) in the presence of this random termination.

The strategy S_l is, of course, a mathematical idealization. It is obvious that in practice Δt cannot be very small because each preventive switching requires some effort. The switching device can also be unreliable, but, unlike the switching upon failure, the state of this device can be checked prior to the preventive switching. Thus, a practical realization of S_l can be formulated in the following way: perform switchings as often as reasonable. Given the corresponding costs and rewards, the problem of obtaining some optimal Δt_{op} can be considered, but this is a topic for future study.

2. SOME GENERALISATIONS

In the same manner a system of n identical components (one operating and $n-1$ standby) can be analyzed. The strategy that maximizes the time to the first failure is that when n-1 equidistant switchings are scheduled. Then, using the same approach, $P_2(S, t_1, t)$, the probability of a system of $n-1$ components performing without failures in $(t_1, t]$, where t_1 is the

time of the first failure, can be maximized etc. This results in the optimal strategy S at each step that can easily be derived. The corresponding l - strategy is obviously characterized by the following sequence of DFs of times between consecutive failures of system's components:

$$F_{1,l}(t) = 1 - \exp\left\{ -\int_0^t \lambda\left(\frac{u}{n}\right) du \right\}; \quad F_{2,l}(t) = 1 - \exp\left\{ -\int_{t_1 - \frac{t_1}{n}}^{t + t_1 - \frac{t_1}{n}} \lambda\left(\frac{u}{n-1}\right) du \right\};$$

$$F_{3,l}(t) = 1 - \exp\left\{ -\int_{t_2 - 2\frac{t_1}{n} - \frac{t_2 - t_1}{n-1}}^{t + t_2 - 2\frac{t_1}{n} - \frac{t_2 - t_1}{n-1}} \lambda\left(\frac{u}{n-2}\right) du \right\}; \dots \tag{10}$$

where $t_1, t_2, \dots$ are the times of the first failure of a component, the second failure of a component,... respectively. Relations (10) follow from the fact that $t_1 - t_1/n$ is "the recalculated starting age" of the integrated system immediately after the first failure, $t_2 - 2t_1/n - (t_2 - t_1)/(n-1)$ is "the recalculated starting age" of the integrated system immediately after the second failure of a component etc (we are excluding the operating times of the failed components). Eventually the concept close to that of minimal repair is reached [2]. It should be noted that this procedure is optimal in the above sense only up to and including the $n-1$th failure. In other words: the optimal strategy increases the times to the first, second,... $n-1$th component failures by decreasing the time between the $n-1$th and the last component failure.

2.1   Consider now the following system of three identical non-repairable components with increasing failure rates. Two components in series start functioning at $t = 0$, while the third one is in the "cold" standby state. After the failure of an operating component, the standby starts operating immediately. The switching from any operating component to the standby can be performed instantaneously at any moment in time. The problem is to find the switching strategy that, under given assumptions, maximizes the probability of the system performing without failure in a given time interval. It is important to note that, in contrast to the setting of Section 1, we are now looking at the total failure of the system: while the last two components are operating (there is no remaining standby component), one of them fails. The strategy, which maximizes $P(S,t)$, the probability of the system performing without failure in $(0,t]$, is of interest. But first, as previously, we shall analyze $P_1(S,t)$, the probability of the system performing without failure of components in $(0,t]$ under an arbitrary switching strategy S.

Let each component operate in $(0, t]$ during intervals of length a_1, a_2 and $a_3 = 2t - a_1 - a_2$, respectively. Consider

$$\max_{a_1, a_2} \left\{ \exp\left(-\int_0^{a_1} \lambda(u)du \right) \exp\left(-\int_0^{a_2} \lambda(u)du \right) \exp\left(-\int_0^{2t-a_1-a_2} \lambda(u)du \right) \right\},$$

which is equivalent to

$$\min_{a_1, a_2} \left\{ \int_0^{a_1} \lambda(u)du + \int_0^{a_2} \lambda(u)du + \int_0^{2t-a_1-a_2} \lambda(u)du \right\}, \tag{11}$$

and finally to the simultaneous equations

$$\begin{cases} \lambda(a_1) - \lambda(2t - a_1 - a_2) = 0 \\ \lambda(a_2) - \lambda(2t - a_1 - a_2) = 0. \end{cases}$$

This system has an obvious unique solution (since $\lambda(t)$ is increasing):

$$a_1 = a_2 = a_3 = \frac{2t}{3}. \tag{12}$$

The corresponding S_l for this case is defined in the following way. Let the first and the second components function in $(0, \Delta t]$, then the second and third components in $(\Delta t, 2\Delta t]$, then the first and third components in $(2\Delta t, 3\Delta t]$ and so on. The last step in constructing this strategy is to perform the limit transition as $\Delta t \to 0$. It is clear that similarly to (9):

$$\max_S T_1(S) = \max_S \int_0^\infty P_1(S, u)du = \int_0^\infty \exp\left\{ -\int_0^t 2\lambda\left(\frac{2u}{3} \right) du \right\} dt \tag{13}$$

Thus the l − strategy is trivially optimal for achieving $\max_S P_1(S, t)$ and is unique for obtaining $\max_S T_1(S)$.

Assume now that the failure, say of the third component, occurred in $(x, x + dx]$ under some arbitrary strategy and that the total functioning time of this component before the failure: $a_3 = 2x - a_1 - a_2; 0 < x < t$ is fixed for this realization. Values of a_1, a_2 denote the total functioning time of the first and the second components, respectively, in $(0, x]$. This means that $a_1 + a_2 \equiv \theta$ is fixed. We want to find the **conditional strategy**, which maximizes the probability of the system functioning without failure in $(0, t]$, given a_3. It is clear that:

$$\max_{a_1, a_2} \lambda(a_3) \exp\left(-\int_0^{a_3} \lambda(u)du \right) da_3 \exp\left(-\int_0^{a_1+t-x} \lambda(u)du \right) \exp\left(-\int_0^{a_2+t-x} \lambda(u)du \right)$$

$$\Rightarrow \max_{a_1}\left\{\exp\left(-\int_0^{a_1+t-x}\lambda(u)du\right)\exp\left(-\int_0^{\theta-a_1+t-x}\lambda(u)du\right)\right\}$$

$$\Rightarrow \min_{a_1}\left\{\int_0^{a_1+t-x}\lambda(u)du+\int_0^{\theta-a_1+t-x}\lambda(u)du\right\}.$$

Thus, as in (3):

$$\lambda(a_1+t-x)-\lambda(\theta-a_1+t-x)=0$$

and finally:

$$a_1=a_2=\frac{\theta}{2}. \tag{14}$$

This result is intuitively obvious, because, given a_3, we can only define an optimal strategy for the first and second components in $(0,x]$, which leads to (14). In reality, x (or a_3) is random and as a result only the $l-$strategy is optimal for the two components over all realizations. What is more, we have arbitrarily assumed that the third component failed first, but it could have been any of them. Hence, for the three components starting at $t=0$ and operating up to the first component failure, only S_l can "service" all situations, being the unique optimal strategy for the case under consideration. It is worthwhile mentioning that, as a result of (12), S_l is trivially optimal for realizations, which have no component failures in $(0,t]$.

Actually, this strategy performs the following operation with our system of increasing failure rate components: At every instant in time it asymptotically ($\Delta t\to 0$) tends to "place" the best components in the operational state, while after the system's failure the remaining non-failed component is more "worn-out" under this strategy than under any other. The best component is defined as the one having the minimum current failure rate. Thus the principle "**the best component should function first**" is realized. We shall come back to discussing this principle while considering some examples.

The above approach can easily be generalized to a system of k identical components (with increasing failure rates) in series, which starts operating, and has n standby components. The corresponding $l-$strategy before the first failure is defined by $k\lambda\left(\frac{k}{k+n}t\right)$, the failure rate of the "integrated component", and then relations similar to (10) can be used. This strategy is

optimal, maximizing the probability of failure-free performance of the system in a given interval of time as well as the mean time to failure.

3. EXAMPLES

3.1 Consider first, as in Section 1, a system of two identical components (main and standby) with exponential DF:

$$F_g(t) = 1 - \exp\{-\lambda t\}. \tag{15}$$

Assume that each component of this system can be instantly repaired upon failure, but the number of repairs is bounded by $m \geq 0$. A situation of this kind often happens in practice. The "total failure" of each component occurs when m repairs have already been performed and the component fails again. Thus, the notions of "failure" and "total failure" differ. The DF of time to total failure follows the Erlangian pattern:

$$F(t) = 1 - \exp\{-\lambda t\} \sum_0^m \frac{(\lambda t)^i}{i!} \tag{16}$$

with increasing failure rate [3]:

$$\lambda(t) = \frac{\lambda \exp\{-\lambda t\}(\lambda t)^m / m!}{\exp\left\{-\lambda t \sum_0^m \frac{(\lambda t)^i}{i!}\right\}}.$$

What strategy should be used to maximize the probability of the system functioning without total failure of any component in the given interval of time $(0, t]$? In Section 1 this probability was denoted by $P_1(S, t)$. A formal answer, based on the previous results, is that switching only at $a = t/2$ (and using the S_l strategy for maximizing T_1) should be used for this purpose. Previously the distribution $F(t)$ was a so-called "black box" DF [2] while now (16) has a concrete form. But the main fact is that now we are able to perform the dynamic strategy of switching, based on the information at hand. This information is just the number of repairs left for each component at any instant of time. From a simple probabilistic reasoning it follows that an optimal strategy would be any strategy S_{opt} that leaves the remaining component in the state with no repairs left after the total failure of the other component. Indeed, it was stated earlier that for this case the switching could not change the probability of the system performing without failure in $(0, t]$ (the failure of the system is defined as the total failure of

both components). In other words: the total number of cycles till the system's failure is $2m+2$ where each cycle has DF (15). Thus, for any optimal strategy of the described type:

$$P_1(S_{opt},t) = \exp\{-\lambda t\}\sum_0^{2m+1}\frac{(\lambda t)^i}{i!} \qquad (17)$$

and for any other, including the one with switching at $a=t/2$ and S_l:

$$P_1(S,t) = \exp\{-\lambda t\}\sum_0^{k}\frac{(\lambda t)^i}{i!} \qquad (18)$$

where $k < 2m+1$ (k in (18) can be random, in which case $P_1(s,t)$ should be understood as the mathematical expectation of the right hand side).

From (17) and (18):

$$P_1(S_{opt},t) > P_1(S,t).$$

This inequality holds for $\forall t \in (0,\infty)$ which is to say that the time to the first failure is stochastically larger (see later) under S_{opt} than under the other strategy. Hence, the optimal black box strategy with switching at $a=t/2$ (and S_l) is worse than the dynamic optimal strategy based on information. The simplest S_{opt} for the case under consideration is the following strategy: the first component starts operating and is replaced by the standby only after the mth failure (it is instantly repaired and there are no more repairs left). The former standby operates till its total failure. It is clear that this strategy maximizes T_1 as well.

3.2   Consider now the system of three identical components described in Section 2, each component satisfying assumptions (15) and (16), two of which are in series and one on standby. We are looking at the total failure of the system: this occurs when two components are operating (there is no remaining standby component) and one of them fails and has no more repairs left. It was proved that the S_l strategy of Section 2 is an optimal black box strategy for this case as well (since $\lambda(t)$ is increasing). We shall define an optimal **information based** strategy of switching for the described system. It is clear that the goal of this strategy is to decrease r, the random number of repairs left ($0 \le r \le m$) for the remaining component after the system has failed. Because the total number of possible of repairs is fixed ($3m$), the random number of cycles n (a cycle is defined as the period between successive failures governed by an exponential DF with failure rate 2λ) will be maximal in some sense and this will lead to the desired result. Indeed, it is clear that $n-1+r=3m$. Denote by r_o and r_a the

number of repairs left for the remaining component under the optimal (to be defined) strategy and under some arbitrary strategy, respectively. Define an optimal strategy S_{opt} as the strategy in accordance with which the best component should function first. The best component is defined as the one having the most repairs left. Thus under the optimal strategy, each time a failure of the component occurs, it is replaced by the standby, if its remaining number of repairs is more than the number that remains after the failure (and instantaneous repair) of the operating component. It is clear that this strategy achieves the following inequality:

$$r_o \leq r_a, \tag{19}$$

which should be understood [4,5] in terms of stochastic ordering (stochastically smaller) i.e:

$$1 - G_o(x) \leq 1 - G_a(x); \quad 0 \leq x \leq m, \tag{20}$$

where $G_o(x)$ and $G_a(x)$ are the discrete DFs of r_o and r_a respectively.

Thus, it follows from above, using the same notation, that:

$$n_o \geq n_a$$

and, eventually

$$P(S_{opt}, t) \geq P(S, t).$$

As previously, inequality (20), means that the expected time to failure of the system under S_{opt} is larger (not smaller) than under the other strategy. Hence, an optimal information based strategy is better than the formal optimal black box strategy S_l, defined in Section 2.

3.3 What is the possible interpretation of the S_l black box principle (the components with lower failure rates should function first) in the situation with information at hand? We shall illustrate this using the simple example given in Section 3.1. There should be no switching at all before the first failure of the operating component, because, due to the exponential DF (15), the components are statistically identical at any instant of time prior to the first failure.

Let equation

$$\lambda_j(t) = \frac{\lambda \exp\{-\lambda t\}(\lambda t)^j / j!}{\exp\left\{-\lambda t \sum_0^j \frac{(\lambda t)^i}{i!}\right\}} \tag{21}$$

define the conditional failure rate $\lambda_j(t)$: the failure rate of the component having $j; 0 \leq j \leq m$ repairs left. It can easily be seen that for any $t \in (0, \infty)$ and $j_1 < j_2$:

$$\lambda_{j_1}(t) > \lambda_{j_2}(t) . \tag{22}$$

Inequality (22) follows from the integral representation of the failure rate for the Gamma (Erlangian) DF [3]:

$$\lambda_j(t)^{-1} = \int_0^\infty \left(1 + \frac{v}{t}\right)^j \exp\{-\lambda v\} dv .$$

Hence, after the first failure we replace the failed component by the standby one, which due to (22), will have a smaller conditional failure rate than the failed component; after its first failure both components will have equal conditional failure rates, etc. It is clear that this strategy is optimal in the sense of equation (17). Actually, it can be defined as the rule: "the best should function first", which was described in Section 3.2. The point is that S_l implements this principle statistically, without additional information (and that is why it is only suboptimal), while the optimal dynamic strategy is information based. The heuristic considerations, presented in Sections 3.2 and 3.3, are intuitively evident (including inequality (19)), but need further mathematical justification in the future.

REFERENCES

[1]   R. BARLOW, F. PROSCHAN, *Statistical Theory of reliability and Life Testing. Probability Models*, Holt, Rinehart and Winston (1975).

[2]   M.S. FINKELSTEIN, *Some notes on two types of minimal repair. Adv. Appl. Prob*. **24**, 226-228 (1992).

[3]   A. HOYLAND, M. RAUSAND, *System Reliability Theory. Models and Statistical Methods*, John Wiley & Sons (1994).

[4]   S.M. ROSS, *Stochastic Processes*. John Wiley & Sons (1996).

[5]   D. STOYAN, *Comparison Methods for Queues and other Stochastic Model,*. John Wiley & Sons (1983).

ACKNOWLEDGEMENTS