

A DECISION SUPPORT TOOL FOR MODELING TIME-TO-EVENT DATA IN POWER SYSTEMS

ABDOLVAHHAB FETANAT

Department of Electrical and Computer Engineering, Behbahan Branch, Islamic Azad University, Behbahan, Iran

E-mail: av-fetanat@behbahaniau.ac.ir

ABSTRACT

Survival analysis, also known as time-to-event analysis, focuses on modeling waiting times until events of interest occur. This paper proposes statistical flowgraph models (SFGMs) as a suitable decision support tool for survival analysis in power systems. SFGMs are useful for modeling time-to-event data that result from a stochastic process. Analysis from the flowgraph model gives an entire waiting time distribution as well as cumulative distribution function, survivor function and hazard function for any total or partial waiting time. The importance of survival analysis in power system is demonstrated when our power supply is disrupted. Two examples illustrated in this paper demonstrate with clarity, the effectiveness of SFGMs to the survival analysis in power systems. One of the advantages of SFGMs is to reduce the multistate models into an equivalent binary-state model. Based on this study, the authors suggest that the use of SFGMs will help decision makers to identify weak points in the system in order to control system conditions.

Keywords: Hazard; Power System; Statistical Flowgraph Model; Survival Analysis; Time-To-Event Data.

1. INTRODUCTION

Survival analysis, also known as failure time analysis or time-to-event analysis, involves the estimation of the distribution of the time it takes for an event to occur to an object depending on its features [1].

The basic function of an electric power system is to supply the load demand as economically as possible within pre-defined continuity, quality and security patterns. However due to the enormous quantity of components in these systems, combined with their unique operation characteristics, there is a possibility of failure of the entire system simply by failure of a crucial or a group of crucial components [2]. This paper describes Statistical flowgraph models for survival analysis in power systems.

Statistical flowgraph models (SFGMs) connect a vast number of areas that are of interest in statistics, mathematics, computer science, and engineering. Some application areas include survival analysis and disease progression in medical studies, reliability engineering, and queuing theory; all of these involve stochastic processes. A statistical flowgraph model is a directed graphical depiction of a finite state stochastic process that is assumed to have the semi-Markov property [3]. They are useful for analyzing time-to-event data. SFGMs are distinct from other graphical models used in

statistics. Directed graphical models use nodes to identify random variables and edges to model causal relationships. SFGMs, on the other hand, use directed graphs to model the outcomes of random variables. In a SFGM, nodes identify the actual physical states of a system; edges model allowable transitions, probabilities of outcomes, and waiting times until the occurrence of outcomes.

SFGMs bring together applied probability techniques such as transforms and approximation methods and meld them with data analysis and statistical methods. They provide a method for accessing the waiting time distribution for any partial or total waiting time [4].

SFGMs were first used in engineering and appeared in the literature of electrical engineering as “signal flow graphs” [5]. Mason [5] was primarily concerned with solving systems of linear equations. Signal flowgraphs are concerned with “transmitting” current with respect to inductance and capacitance. Butler and Huzurbazar [6] adapted SFGMs for use in Bayesian stochastic models. Since then, the use and theory of SFGMs has continued to expand. Huzurbazar [7] used SFGMs to generalize phase-type distributions. Huzurbazar [8] demonstrated a Bayesian application of SFGMs on a complex cellular telephone network. Butler and Huzurbazar [9] improved on some of the techniques used in statistical flowgraph modeling

and demonstrated their use in Bayesian prediction of waiting times in queuing theory. Yau and Huzurbazar [10] show how SFGMs can be used to model incomplete data in multistate systems. The theory linking semi-Markov processes with multistate models using SFGMs was explained in [11]. Huzurbazar [4] provides an excellent example of how Bayesian SFGMs can be applied in various fields, using an example in construction project management. Huzurbazar [12] is a full length text devoted to SFGMs and its applications. Other applications of SFGMs can be found in [13-15].

2. STATISTICAL FLOWGRAPH MODELS

Throughout this paper, we will assume that the time to event T is a continuous random waiting time that is nonnegative.

Actually, the distribution of T has four representations that are equivalent. Each representation is useful in specific contexts. These representations include the probability density function, the survivor function, the cumulative distribution function, and the hazard function. The survivor function and the hazard function are very useful to survival analysis methodology.

Before we describe SFGMs, we present two definitions [12].

Definition (survivor function). In the analysis of time-to-event data, one quantity of interest is the probability that a component survives beyond some time t starting from the initial time $t = 0$. The word *survives* is used quite generally to mean that the event of interest has not yet occurred. Let T be the random waiting time until the event of interest occurs. The quantity $P(T > t)$ represents the probability that the event of interest has not occurred. In survival analysis, this is called the *survivor function* and is denoted by

$$S(t) = P(T > t) \quad (1)$$

For instance, the survivor function gives the probability that a component survives beyond some specified time t . In engineering applications, this is called the *reliability function* and is denoted by $R(t) = P(T > t)$.

Definition (hazard function). Suppose that T is the random waiting time until the event of interest occurs, and let $f(t)$, $F(t)$, and $S(t)$ denote the density, cumulative distribution function (CDF), and survivor function of T . Then the *hazard function* is defined as

$$h(t) = \frac{f(t)}{S(t)} \quad (2)$$

This is also called the *instantaneous failure rate*. The hazard gives the instantaneous rate of failure at time t given that the component survives up to time t . Although the hazard is technically a rate, it is often spoken of in probabilistic terms as the chance that a component “fails” at time t conditional on having survived up to that time. In fact, $h(t)$ is *not* a probability density function and it can take on any value greater than or equal to zero. We can think of the hazard function as an component’s propensity to fail in the next short interval of time, given that the component has survived to time t .

Some common relationships between these functions are

$$f(t) = \frac{dF(t)}{dt} = -S'(t) \quad (3)$$

$$S(t) = 1 - F(t) \quad (4)$$

$$S(t) = \exp \left[- \int_0^t h(u) du \right] \quad (5)$$

$$h(t) = - \frac{d \log S(t)}{dt} \quad (6)$$

Multistate models are used to describe time-to-event data. They model stochastic processes that progress through various stages [14]. Time-to-event data analysis focuses on modeling waiting times until events of interest occur [12]. Today’s complex systems make the analysis of multistate models very important in reliability. Statistical flowgraph models are one type of multistate model. Statistical flowgraphs model potential outcomes, probabilities of outcomes, and waiting times for the outcomes to occur. They can be used to model complex system behavior, time to total or partial system failure, time to repair of components or the entire system, and to predict system reliability. Current methods in reliability for complex systems require that all the component waiting times be from the same distributional family such as the Weibull so that information may be easily combined. In addition, some methods for complex systems such as fault trees restrict themselves to binomial data so that large amounts of information may be easily aggregated. SFGMs allow each component or set of components to have its own distribution and statistical flowgraph algebra provides a way to combine these varied distributions [14].

SFGMs provide a link between traditional multistate models and semi-Markov processes. They model semi-Markov processes and provide a practical alternative methodology for data analysis. They are useful when a continuous time semi-Markov process defines the transition times

between states and interest focuses on estimating the density, CDF, survivor function and hazard function of the process [4]. Semi-Markov analysis is restricted to cases where the waiting times between states all come from the same family of distributions or each transition in the multistate model is modeled by a separate proportional hazards model [12]. Although semi-Markov processes are a natural extension for multistate models, in practice, data analysis for semi-Markov processes can be quite difficult, especially when returns to a state are allowed [16]. SFGMs circumvent this difficulty by working in the moment generating function (MGF) domain.

Statistical flowgraph models provide an extension of traditional multistate models to include semi-Markov processes. They allow vastly different distributions for waiting times between states, and they provide a closed form for the distribution of quantities of interest, such as total or partial waiting time in a stochastic network [12]. In addition to allowing a variety of distributions to be used within the stages of the multistate model, the statistical flowgraph methodology also easily handles reversibility. In the engineering context, this means that a failed component can be repaired [14].

SFGMs have been used in both Bayesian and frequentist frameworks [14]. The final result of a Bayesian SFGM is a posterior predictive density (PPD) of the waiting time from one state to another, and similarly for frequentist SFGMs the result is a probability density function (PDF).

There is a systematic way to implement statistical flowgraphs. The first step is to propose the system diagram or graphical model. We design our graph by identifying the states the process can assume. These states are represented by the nodes of the graph. Next, we identify the possible transitions between states, which are the directed branches (or edges) of the graph. Once the graphical model is in place, we examine the data to suggest appropriate distributions for the branch transition times. The selected distributions model the time it takes to transition from one state to another. Next, we find the MGFs or Laplace Transforms (LTs) of the distributions assigned to the branches of the SFGM [3].

The powerful capability of SFGMs to power system survival analysis is demonstrated via two examples from power system operation and power system protection.

2.1. Example 1: statistical flowgraph models for power system operation

For purpose of analyzing power system security and designing appropriate control systems, it is helpful to conceptually classify the system-operating conditions into five states: normal, alert, emergency, in extremis, and restorative. Fig. 1 depicts these operating states and the ways in which transition can take place from one state to another [17].

Fig. 1 SFGM For Operating States Of A Power System At End Of Text

In the normal state, all system variables are within the normal range and no equipment is being overloaded. The system operates in a secure manner and is able to withstand a contingency without violating any of the constraints. The system enters the alert state if the security level falls below a certain limit of adequacy, or if the possibility of a disturbance increases because of adverse weather conditions such as the approach of severe storms. In this state, all system variables are still within the acceptable range and all constraints are satisfied. However, the system has been weakened to a level where a contingency may cause an overloading of equipment that places the system in an emergency state. If the disturbance is very severe, the in extremis (or extreme emergency) state may result directly from the alert state. Preventive action, such as generation shifting (security dispatch) or increased reserve, can be taken to restore the system to the normal state. If the restorative steps do not succeed, the system remains in the alert states. The system enters the emergency state if a sufficiently severe disturbance occurs when the system is in the alert state. In this state, voltages at many buses are low and/or equipment loadings exceed short-term emergency ratings. The system is still intact and may be restored to the alert state by the initiating of emergency control actions: fault clearing, excitation control, fast-valving, generation tripping, generation run-back, high voltage dc (HVDC) modulation, and load curtailment. If the above measures are not applied or are ineffective, the system is in extremis; the result is cascading outages and possibly a shut-down of a major portion of the system. Control actions, such as load shedding and controlled system separation are aimed at saving as much of the system as possible from a widespread blackout. The restorative state represents a condition in which control action is being taken to reconnect all the facilities and to restore system load. The system transits from this state to either the alert state or the normal state, depending on the system conditions. Characterization of the system conditions into the



five states as described above provides a framework in which control strategies can be developed and operator actions identified to deal effectively with each state. For a system that has been disturbed and that has entered a degraded operating state, power system controls assist the operator in returning the system to a normal state. If the disturbance is small, power system controls by themselves may be able to achieve this task. However, if the disturbance is large, it is possible that operator actions such as generation rescheduling or element switching may be required for a return to the normal state [17].

In a SFGM, the states or nodes represent outcomes. The nodes are connected by directed line segments called branches. These branches are labeled with transmittances. These transmittances are labeled with the “transition probability \times moment generating function of the waiting time distribution in the previous state” which is a quantity called the branch transmittance. The waiting times on the branches can be any parametric distributions that admit MGFs [4]. We use the branch transmittances of a SFGM to solve for the MGF of the distribution of the waiting time of interest. The next section describes how to solve the statistical flowgraph of Fig. 1.

2.1.1. Solving a statistical flowgraph model

The term *solving* a SFGM refers to finding the MGF of the overall waiting time distribution from one state to another.

We use a procedure based on Mason’s rule, described in [5], to solve the SFGM. Mason’s rule did not involve probabilities or MGFs. However, SFGMs can be solved by applying Mason’s rule to the branch transmittances. When the system is certain to pass eventually from state u to state v , the transmittance is the MGF of the waiting time distribution. Practical use of Mason’s rule entails identifying all of the paths from u to v , the loops involved in those paths and then solving for the overall MGF. It requires computing the transmittance for every distinct path from the initial state to the end state and adjusting for the transmittances of various loops. Before we describe Mason’s rule, we present the following definitions [12].

Definition (path). A path is any possible sequence of nodes from an initial state to an end state that does not pass through any intermediate node more than once.

Definition (path transmittance). A path transmittance is the product of all of the branch transmittances for that path.

Definition (loop). A loop is any closed path that returns to the initiating state without passing through any state more than once.

Mason’s rule differentiates among different types of loops. These are defined as follows [12].

Definition (first-order loop). A first-order loop is any closed path that returns to the initiating node without passing through any node more than once.

Definition (second-order loop). A second-order loop consists of two non-touching first-order loops.

Definition (J^{th} -order loop). A J^{th} -order loop consists of J non-touching first-order loops.

Consider the SFGM in Fig. 1. It has five states. Suppose that our interest is in predicting the waiting time distribution from 2 \rightarrow 4. There are two paths from 2 \rightarrow 4: 2 \rightarrow 4 and 2 \rightarrow 3 \rightarrow 4. The transmittances of these paths required for the application of Mason’s rule are given in Table 1. The transmittance of a first-order loop is the product of the individual branch transmittances involved in the path. Table 1 gives the transmittance for this loop. The transmittance of a higher-order loop is the product of the transmittances of the first-order loops it contains. There are no higher-order loops in this SFGM.

Table 1 insert here, exactly.

Let T be the overall waiting time from state u to state v . The general form of Mason’s rule that gives the MGF of the distribution of T , is

$$M(s) = \frac{\sum_I P_I(s)[1 + \sum_J (-1)^J L_J^I(s)]}{1 + \sum_J (-1)^J L_J(s)} \tag{7}$$

where:

$P_I(s)$ is the transmittance for the I^{th} path.

$L_J(s)$ in the denominator is the sum of the transmittances over the J^{th} -order loops.

$L_J^I(s)$ is the sum of the transmittances over the J^{th} -order loops sharing no common nodes with the I^{th} path (i.e., loops not touching that path).

Eq. (7) for the MGF of the distribution T , the waiting time from 2 to 4, in Fig. 1 is

$$M(s) = \frac{p_{24}M_{24}(s) + p_{23}p_{34}M_{23}(s)M_{34}(s)}{1 - L_1(s)} \tag{8}$$

Where

$$L_1(s) = p_{12}p_{21}M_{12}(s)M_{21}(s) + p_{23}p_{32}M_{23}(s)M_{32}(s) + p_{24}p_{45}p_{52}M_{24}(s)M_{45}(s)M_{52}(s) + p_{23}p_{34}p_{45}p_{52}M_{23}(s)M_{34}(s)M_{45}(s)M_{52}(s) + p_{12}p_{24}p_{45}p_{51}M_{12}(s)M_{24}(s)M_{45}(s)M_{51}(s) + p_{12}p_{23}p_{34}p_{45}p_{51}M_{12}(s)M_{23}(s)M_{34}(s)M_{45}(s)M_{51}(s)$$

and $L_J^I(s) = 0$, since all loops touch the path.



We assume a variety of waiting time distributions such as Inverse Gaussian (λ_1, λ_2) , Gamma (α, β) and Exponential (μ) given in Table 2. If we had data on the real system, we would use that to select candidate distributions [4]. Parameter estimation is done by standard methods such as maximum likelihood [4] or the entire problem can be solved from the viewpoint of prediction using Bayesian methods as in Section 2.1.3.

For now, we assume the following parameter values

$$p_{12} = 0.6, p_{21} = 0.3, p_{23} = 0.7, p_{24} = 1 - p_{21} - p_{23}, p_{32} = 0.5, p_{34} = 1 - p_{32}, p_{45} = 0.4, p_{51} = 0.6, p_{52} = 1 - p_{51}, \lambda_{1,12} = 20, \lambda_{2,12} = 5, \mu_{21} = 0.3, \alpha_{23} = 3, \beta_{23} = 4, \mu_{24} = 1, \mu_{32} = 3, \mu_{34} = 1, \alpha_{45} = 4, \beta_{45} = 1, \mu_{52} = 0.3, \mu_{51} = 0.5.$$

Table 2 Waiting Time Distributions For The Statistical Flowgraph Of Fig. 1 At End Of Text.

The MGF for the overall waiting time using these distributions in Eq. (8) is (from MAPLE software)

$$M(s) = \left(\left(p_{23} \left(\frac{\beta_{23}}{\beta_{23}-s} \right)^{\alpha_{23}} p_{34} \mu_{34} (\mu_{34} - s)^{-1} + (p_{24} \mu_{24} (\mu_{24} - s)^{-1}) \right) \times \left(1 - \left(p_{12} e^{\lambda_{1,12} \lambda_{2,12} - (\lambda_{1,12}^2 (\lambda_{2,12}^2 - 2s))^{0.5}} p_{21} \mu_{21} (\mu_{21} - s)^{-1} \right) - \left(p_{12} e^{\lambda_{1,12} \lambda_{2,12} - (\lambda_{1,12}^2 (\lambda_{2,12}^2 - 2s))^{0.5}} p_{23} \left(\frac{\beta_{23}}{\beta_{23}-s} \right)^{\alpha_{23}} p_{34} \mu_{34} p_{45} \times \left(\frac{\beta_{45}}{\beta_{45}-s} \right)^{\alpha_{45}} p_{51} \mu_{51} (\mu_{34} - s)^{-1} (\mu_{51} - s)^{-1} \right) - \left(p_{12} e^{\lambda_{1,12} \lambda_{2,12} - (\lambda_{1,12}^2 (\lambda_{2,12}^2 - 2s))^{0.5}} p_{24} \mu_{24} p_{45} \right) \times \left(\frac{\beta_{45}}{\beta_{45}-s} \right)^{\alpha_{45}} p_{51} \mu_{51} (\mu_{24} - s)^{-1} (\mu_{51} - s)^{-1} \right) - \left(p_{23} \left(\frac{\beta_{23}}{\beta_{23}-s} \right)^{\alpha_{23}} p_{32} \mu_{32} (\mu_{32} - s)^{-1} \right) - \left(p_{23} \left(\frac{\beta_{23}}{\beta_{23}-s} \right)^{\alpha_{23}} p_{34} \mu_{34} p_{45} \left(\frac{\beta_{45}}{\beta_{45}-s} \right)^{\alpha_{45}} p_{52} \mu_{52} \right) (\mu_{34} - s)^{-1} (\mu_{52} - s)^{-1} \right) - \left(p_{24} \mu_{24} p_{45} \left(\frac{\beta_{45}}{\beta_{45}-s} \right)^{\alpha_{45}} p_{52} \mu_{52} (\mu_{24} - s)^{-1} (\mu_{52} - s)^{-1} \right) \right)^{-1}$$

We could replace Fig. 1 with an equivalent SFGM that has only two states 2 and 4, with one connecting branch. The MGF assigned to the branch is $M(s)$ as defined in Eq.(9).

Solving SFGMs gives us the MGF of the overall waiting time distribution of interest; however, we still have not obtained the distribution itself. The next section discusses the conversion of MGFs such as Eq.(9) to waiting time density.

2.1.2. Inversion of statistical flowgraph MGF

Note that our interest with SFGMs is to convert the resulting MGF into a density, CDF, survivor function and hazard function.

Parametric SFGMs have used MGF to represent the probability of transition between states. This has greatly limited the number of distributions available for SFGMs, since not all distributions have MGFs. Popular distributions such as the lognormal or certain Weibulls do not have MGFs and have not been used in SFGMs. By using complex Laplace Transforms (LTs) in lieu of MGFs we can use all continuous and differentiable parametric distributions in SFGMs. Complex LTs are a generalization of MGFs and characteristic functions and exist for all lifetime distributions. The Laplace Transform, as defined in statistics, is a simplification of the LT as defined in mathematics. In mathematics, the Laplace Transform, $L(z)$, is defined for all real and complex z , while in statistics it is usually only defined for real z . Therefore, we use the term complex LT to avoid ambiguity [3]. A complex LT for a positive random variable T , defined on $[0, \infty)$ with

$$Z = x + iy \text{ and } x \geq 0, \text{ is } L(z) = \int_0^\infty e^{-zt} f(t) dt \tag{10}$$

If the MGF has a closed analytic form, then so does the complex LT, and finding the complex LT for the random variable T is a transformation similar to finding the MGF. However, for distributions that do not have closed form MGFs we must find the complex LT or MGF by numerical integration. Examples of distributions that have closed form complex LTs are the gamma and inverse Gaussian distributions, and examples that do not are the lognormal and Weibull distributions. Using the complex LT does not change any theory we introduced in Section 2.1.1. Mason's rule still applies; all we do is replace the MGFs with their associated complex LTs [3].

The most common method for converting the MGF of the overall waiting time of interest into an approximate density is the saddlepoint method. This method is cumbersome when solving complicated SFGMs. It cannot handle distributions that do not have MGFs. Another possible method of inversion is the EULER method, developed by Abate and Whitt [18]. The EULER method requires that the densities involved be continuous and differentiable. Thus by using complex LTs in the place of MGFs on the branches of a SFGM, we can use any relatively continuous and differentiable



distribution in statistical flowgraph modeling. This is a big step forward for SFGMs [3]. The Euler method uses the Bromwich integral and Euler summation (and hence its name). The Bromwich contour inversion integral is [3]:

$$f(t) = \frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} e^{zt} L(z) dz, \quad (11)$$

Where $i = \sqrt{-1}$, $L(z)$ is the complex LT, and the contour is any vertical line $z = a$ such that $L(a)$ has no singularities on or to the right of it. With a change of variable and some manipulation, $f(t)$ can be rewritten as

$$f(t) = \frac{2e^{at}}{\pi} \int_0^\infty Re[L(a + iu)] \cos(ut) du, \quad (12)$$

Where $Re[z]$ is the real part of a complex number. Using the trapezoidal rule with step size $\pi/(2t)$ and letting $a = A/(2t)$ gives the approximation

$$f(t) \approx \frac{e^{A/2}}{2t} Re \left[L \left(\frac{A}{2t} \right) \right] + \frac{e^{A/2}}{t} \sum_{k=1}^{\infty} (-1)^k Re \left[L \left(\frac{A+2k\pi i}{2t} \right) \right]. \quad (13)$$

This is a nearly alternating series so Abate and Whitt [18] use Euler summation as an acceleration method. Combining these, our approximation for the density of one random variable T , is

$$f(t) \approx \sum_{j=0}^m \left(\frac{e^{A/2}}{t} \frac{2^{-m} m!}{j!(m-j)!} \left(\frac{Re \left[L \left(\frac{A}{2t} \right) \right]}{2} + \sum_{k=1}^{n+j} (-1)^k Re \left[L \left(\frac{A+2k\pi i}{2t} \right) \right] \right) \right), \quad (14)$$

or,

$$f(t) \approx \frac{e^{A/2}}{t} \sum_{k=0}^{m+n} (-1)^k w_k Re \left[L \left(\frac{A+2k\pi i}{2t} \right) \right], \quad (15)$$

Where

$$w_0 = \frac{1}{2}, w_k = 1 \text{ for } k = 1 \dots n,$$

and

$$w_k = w_{k-1} - \frac{2^{-m} m!}{(k-n-1)! (m+n+1-k)!} \text{ for } k = n+1 \dots n+m.$$

Abate and Whitt [19] recommend setting $m = 11, n = 15$, and increasing n if better accuracy is required. This approximation contains two different errors. First is the error introduced by the trapezoidal approximation, and the second by the truncated sum and Euler acceleration. Abate and Whitt [19] show how to bound the first type of error by choosing A (often they choose $A = 18.4$). However, the error introduced by the truncated sum

and Euler acceleration cannot be bounded, only estimated.

All of the distributions used in this example have closed form MGFs. Now, we use the Euler method to convert the MGF of the overall waiting time from 2 to 4 into a density. The resulting density, CDF, etc., are given in Fig. 2 (from MATLAB).

Fig. 2 Density Function, CDF, Survivor Function And Hazard Function For Example 1 At End Of Text.

SFGMs focus on modeling the observed state-to-state waiting times. This type of modeling allows us to realistically model the shape of the hazard. The shape of the hazard is increasing and decreasing at varying rates. This is typical of hazards from SFGMs since we are modeling the actual waiting times and not making assumptions about the hazards. Based on these, the purpose of plots presented in Fig. 2 is to help energy managers better understand the hazards and opportunities they face, to evaluate the options available for their control and finally to select optimal energy policies. SFGM is a useful decision support tool for survival analysis in power systems. The term decision support (DS) contains the word *support*, which refers to supporting people in making decisions. Thus, DS is concerned with human decision making, especially in terms of helping people improving their decision making.

Analysis from the flowgraph model gives an entire waiting time distribution as well as the system reliability (survivor) and hazard functions for any total or partial waiting time. For decision making, this immediately provides the mean, median, or any percentile of the waiting time distribution for a variety of input distributions [12].

In addition to the functions $S(t)$ and $h(t)$, a common metric in power system is the mean time to failure (MTTF) which is the average time to the first failure. It can be obtained from the mean of the probability density of the time to failure $f(t)$:

$$MTTF = \int_0^\infty t f(t) dt \quad (16)$$

If $S(t)$ is the system reliability (survivor) function, the integral in Eq. (16) becomes $MTTF = -\int_0^\infty t dS(t)$ which, after integrating by parts [20-22], gives

$$MTTF = \int_0^\infty S(t) dt \quad (17)$$



2.1.3. Bayesian statistical flowgraph models

Bayesian SFGMs use the same principles as frequentist SFGMs. The primary difference is that the Bayesian framework provides a posterior predictive distribution (PPD) as the final result. This is very useful since prediction is the primary focus of SFGMs. This section describes the computation of Bayesian predictive density, cumulative distribution function, survivor function, and hazard function for SFGMs using Euler method and slice sampling technique. Information about the process allows us to solve the SFGM as seen in Section 2.1.1 and Section 2.1.2. Observations on the various transitions allow us to consider parametric models for the branch waiting times of the SFGM. Our interest lies in making inferences about the unknown parameters in the model, updating our knowledge of the mechanism that generated the data, evaluating branch models, and ultimately using our data and the SFGM for prediction. Bayesian analysis formally incorporates subjective information about a problem into the analysis.

We can think about our model as the theory that drives what we can observe. For a given parameter vector θ , we expect things to behave in a certain way. We use the data to give us additional information about θ , but we can never know θ . We know that our theory (model) is incorrect, so the question then becomes: How useful is it? What is often important is whether we can use the formulated model to make useful predictions about new observations, $Z = (Z_1, \dots, Z_h)$, that are conditionally independent of \mathcal{D} , the observed data, given θ . In a flowgraph model, these future observations will generally be total or partial waiting times [12].

Definition (*Bayes theorem*). The posterior distribution of parameter vector θ given the data \mathcal{D} is defined by

$$\pi(\theta|\mathcal{D}) \propto \mathcal{L}(\theta|\mathcal{D})\pi(\theta) \tag{18}$$

where $\mathcal{L}(\theta|\mathcal{D})$ is the likelihood function and $\pi(\theta)$ is the prior [12].

The Bayesian predictive density of a future observable Z given data \mathcal{D} about a system with parameter vector θ is

$$f_Z(z|\mathcal{D}) = \frac{\int f_Z(z|\theta)\mathcal{L}(\theta|\mathcal{D})\pi(\theta)d\theta}{\int \mathcal{L}(\theta|\mathcal{D})\pi(\theta)d\theta} \tag{19}$$

where Z has density $f_Z(z|\theta)$ [12].

In general, we do not know $f_Z(z|\theta)$ in Eq.(19). In this case, we use $\tilde{f}_Z(z|\theta)$ in place of $f_Z(z|\theta)$. $\tilde{f}_Z(z|\theta)$ is computed using the Euler method i.e. Eq. (14) or Eq.(15). Practically, this involves generating a random sample of θ values from the posterior and performing the following computations for each θ value [12]:

1. Convert the statistical flowgraph MGF, $M_{Z|\theta}(s)$, to a density using the Euler method for each value z of Z at which the density is desired.
2. Treat each density as a vector of values (evaluated at each value z in computation 1) indexed by θ_r .
3. For each z value, average the densities over the θ values to give

$$\hat{f}_Z(z|\mathcal{D}) = \frac{\sum_{r=1}^M \tilde{f}_Z(z|\theta_r)}{M} \tag{20}$$

where $\tilde{f}_Z(z|\theta_r)$ is the density derived from the Euler method and M is the number of samples. For large M , this sum approximates the integral Eq.(19).

There are many methods for sampling from the posterior distribution (e.g. rejection sampling, importance sampling, Gibbs sampling, etc. [12]), and some methods are better suited than others for specific problems. A recently developed technique that lends itself to sampling from the posterior with SFGMs is a Markov Chain Monte Carlo (MCMC) method called slice sampling [12]. Slice sampling was introduced by Neal [23] and is presented in detail in [24]. The basic premise of slice sampling is that one can sample from any univariate distribution by sampling points uniformly under the graph of its density function. Multivariate slice sampling can be performed by applying a single variable slice sampling method to each variable in sequence [12]. In this study, we use Slice sampling for sampling from the posterior distribution.

Our goal is to compute the Bayesian predictive density of the overall waiting time from $2 \rightarrow 4$ using Eq.(20). We let T be this random time. Our parameter vector θ is $\theta = (p_{12}, p_{21}, p_{23}, p_{24}, p_{32}, p_{34}, p_{45}, p_{51}, p_{52}, \lambda_{1,12}, \lambda_{2,12}, \mu_{21}, \alpha_{23}, \beta_{23}, \mu_{24}, \mu_{32}, \mu_{34}, \alpha_{45}, \beta_{45}, \mu_{52}, \mu_{51})$. We assume Unif(0,1) priors on $p_{12}, p_{21}, p_{23}, p_{24}, p_{32}, p_{34}, p_{45}, p_{51}$ and p_{52} . The remaining priors are



$$\begin{aligned}
 \pi(\mu_{24}), \pi(\mu_{34}), \pi(\alpha_{45}), \pi(\beta_{45}) &\sim \text{Unif}(0,4) \\
 \pi(\mu_{21}), \pi(\mu_{52}), \pi(\mu_{51}) &\sim \text{Unif}(0,1) \\
 \pi(\alpha_{23}), \pi(\beta_{23}) &\sim \text{Unif}(0,2) \\
 \pi(\lambda_{112}) &\sim \text{Unif}(1,5) \\
 \pi(\lambda_{212}) &\sim \text{Unif}(1,10) \\
 \pi(\mu_{32}) &\sim \text{Beta}(2,1)
 \end{aligned} \tag{21}$$

This computation, using $M = 1000$ values, gives the Bayes predictive density, CDF, survivor function, and hazard function of the waiting time of interest in Fig.3. Note that we never actually get the exact predictive density; we estimate it via the random samples.

Fig. 3. Bayes Predictive Density Function, CDF, Survivor Function And Hazard Function For Example 1 At End Of Text.

2.2. Example 2: statistical flowgraph models for power system protection: protective relay testing

The goal of protective relay testing is to maximize the availability of protection and minimize risk of relay misoperation. Having defined the necessary tests and monitoring methods for protective relays, it is now necessary to optimize the testing interval [25].

Schweitzer [25] introduced a nine-state model defined by the operating condition of the relay and the protected component. The model accounts for relay self-testing, but does not account for other monitoring means. Fig. 4 shows a ten-state model that accounts for self-testing and models routine relay verification through other simple checks. The circles represent the model states. The arrows represent the transition paths between the states.

Fig. 4 SFGM For A Protection/Component System At End Of Text.

The model assumes that when a fault occurs while the relay is out of service, a larger portion of the power system is isolated than was actually necessary to remove the fault. The relay could be out of service because of a failure, testing, or repairs.

The probability model is divided into four quadrants representing the condition of the relay (Protection (P)) and the line (Component (C)). State 1 represents a normal operating condition where the line is energized (C UP) and the relay is operating properly (P UP). When a line fault occurs, the Component makes the transition to a down (DN) state, represented by state 2. In state 2, the line is faulted and the relay signals the circuit breaker to trip. Circuit breaker operation takes the

model system to state 6, where the line is isolated (ISO). The line is repaired and re-energized, taking the model back to state 1.

States 5, 3, 9 and 10 represent conditions where the relay is out of service and unavailable to trip if a fault occurs. In state 5, the relay is out of service for routine testing. In states 3, 9, and 10, the relay is out of service due to a relay failure. State 9 represents the relay under repair (REP). The model enters state 9 from state 1 when a relay self-test detects a failure. The model system enters state 9 from state 3 when a routine test detects a relay failure. The model enters state 9 from state 10 when a meter check detects a failure.

The model enters state 3 from state 1 when a relay failure occurs that is not detected by the relay self-test function and could not be detected by meter checks.

A failure not detected by either self-tests or meter tests are only detected by routine testing or by observing a misoperation. The model enters state 4 if a fault occurs while the relay is out of service, or if a common-cause failure of the relay and power system occurs. If a fault occurs while the relay is out of service, remote backup protection operates to isolate the fault. When the remote protection operates, a larger portion of the power system is taken out of service than would have been if the failed relay had operated properly. This is represented in state 4 and state 8 by the isolation of C and X, where X is the additional equipment removed from service by the backup relay trip operation.

Consider the SFGM in Fig. 4. Quantity of interest includes predicting the distribution of the total waiting time from state 1 to state 4. There are eight paths from 1 to 4 (e.g.1 → 3 → 9 → 4). Also, there are nineteen first-order loops (e.g.1 → 2 → 6 → 7 → 9 → 1) and six second-order loops (e.g. 1 → 3 → 9 → 1 with 6 → 7 → 6). In this analysis, the waiting time distribution for transition from any state u to any state v is assumed to be Exponential (μ). Applying Mason's rule we find the MGF of the overall waiting time distribution from state 1 to state 4. Finally, we transform this MGF into a density, CDF, etc. They are shown in Fig. 5.

Fig. 5 Density Function, CDF, Survivor Function And Hazard Function For Example 2.

To compute the Bayes predictive density of the total waiting time distribution from 1 to 4, we assume Unif(0,1) priors on all of the parameters. Fig. 6 shows the Bayes predictive density, CDF, etc., for this illustration.

Fig. 6 Bayes Predictive Density Function, CDF, Survivor Function And Hazard Function For Example 2 At End Of Text

3. CONCLUSION

The aim of survival analysis is to estimate the distribution of the time it takes for events to occur. This paper proposed statistical flowgraph models (SFGMs) as a new approach to do survival analysis in power systems. This decision support tool (using the semi-Markov assumption) provides a way for analyzing time-to-event data and constructing corresponding Bayesian predictive distributions. SFGMs allow data analysis for semi-Markov processes using non-exponential waiting times. They can be applied without becoming intimately familiar with the mathematical theory of stochastic processes. While the increase number of states improves the accuracy of modeling power systems, it also brings the problem of whether proposed multistate models can be compatible with practical studies. Therefore, it is necessary to find a method to reduce multistate models into an equivalent binary-state model, to ensure proposed models are practical. SFGMs provide a systematic way that can be used to reduce complex power systems models. The approach advances the state of the art in power system survival analysis. It also lays a general foundation for effective and systematic power system management.

SFGMs model the observable waiting times rather than the hazards and as such, they do not directly make any assumptions about the shape of the hazard. For decision making, this immediately provides the mean, median, or any percentile of the waiting time distribution for a variety of input distributions [12].

ACKNOWLEDGMENTS

I am thankful to Department of Electrical and Computer Engineering, Behbahan Branch, Islamic Azad University, Behbahan, Iran for their support and collaboration on the design of the presented exercises and sharing their pandemic preparedness plans with me.

REFERENCES:

- [1] D. G. Kleinbaum, M. Klein, *Survival Analysis: A Self-Learning Text*, second ed., Springer-Verlag, New York, 2005.
- [2] Leonel de Magalhães Carvalho, *Using evolutionary swarms (epso) in power system reliability indices calculation*, Master thesis, University of Porto, Porto, Portugal, 2008.
- [3] R.L. Warr, *Generalizations of the statistical flowgraph model framework*, PhD thesis, University of New Mexico, Albuquerque, NM, 2010.
- [4] A.V. Huzurbazar, *Flowgraph models: a Bayesian case study in construction engineering*, *J. Stat. Plann. Infer.* 129 (2005) 181–193.
- [5] S.J. Mason, *Feedback theory: some properties of signal flow graphs*, *Proc. Inst. Radio Eng.* 41(1953) 1144–1156.
- [6] R.W. Butler, A.V. Huzurbazar, *Stochastic network models for survival analysis*, *J. Am. Stat. Assoc.* 92 (1997) 246–257.
- [7] A.V. Huzurbazar, *Flowgraph models for generalized phase type distributions with non-exponential waiting times*, *Scand. J. Stat.* 26 (1999) 145–157.
- [8] A.V. Huzurbazar, *Modeling and analysis of engineering systems data using flowgraph models*, *Technometrics*, 42 (2000) 300–306.
- [9] R.W. Butler, A.V. Huzurbazar, *Bayesian prediction of waiting times in stochastic models*, *Can. J. Stat.* 28 (2000) 311–325.
- [10] C. L. Yau, A.V. Huzurbazar, *Analysis of censored and incomplete survival data using flowgraph models*, *Stat. Med.* 21 (2002) 3727–3743.
- [11] A.V. Huzurbazar, *Multistate models, flowgraph models, and semi-Markov Processes*, *Comm. Stat. Theor. Meth.* 33 (2004) 457–474.
- [12] A.V. Huzurbazar, *Flowgraph models for multistate time-to-event data*, John Wiley & Sons, New York, 2005.
- [13] A.V. Huzurbazar, *Modeling time-to-event data using flowgraph models*, in: N. Balakrishnan (ed.), *Advances on methodological and applied aspects of probability and statistics*, Taylor and Francis, New York, 2000, pp. 561–571.
- [14] A.V. Huzurbazar, B.J. Williams, *Flowgraph Models for complex multistate system reliability*, in: A.G. Wilson, N. Limnios, S.A. Keller-McNulty, Y.M. Armijo (eds.), *Modern statistical and mathematical methods in reliability*, Singapore: World Scientific, 2005, pp. 247–262.
- [15] A.V. Huzurbazar, *Modelling survival data using flowgraph models*, in: N. Balakrishnan, C.R. Rao (eds.), *Handbook of statistics volume*



- 23: advances in survival analysis, Elsevier, Amsterdam, 2004, pp. 729–746.
- [16] P. K. Andersen, N. Keiding, Multi-state models for event history analysis, *Stat. Methods Med. Res.* 11 (2002) 91–115.
- [17] P. Kundur, *Power system stability and control*, McGraw-Hill, 1994.
- [18] J. Abate, W. Whitt, The Fourier-series method for inverting transforms of probability distributions, *Queueing Syst.* 10 (1992) 5–88.
- [19] J. Abate, W. Whitt, Numerical inversion of Laplace transforms of probability distributions, *Inform. J. Comput.* 7 (1995) 36–43.
- [20] D.L. Grosh, *A Primer of Reliability Theory*, John Wiley & Sons, Inc., New York, 1989.
- [21] R. Billinton, R. N. Allan, *Reliability Evaluation of Engineering Systems: Concepts and Techniques*, Plenum Press, first ed., New York, 1983.
- [22] J. Endrenyi, *Reliability modeling in electric power systems*, John Wiley & Sons, New York, 1978.
- [23] R. Neal, Markov chain Monte Carlo methods based on “slicing” the density function, Technical Report 9722, Department of Statistics, University of Toronto, Ontario, Canada, 1997.
- [24] R. Neal, Slice sampling, *Ann. Stat.* 31 (2003) 705–767.
- [25] E.O. Schweitzer III, J.J. Kumm, D. Hou, Assessing the effectiveness of self-tests and other monitoring means in protective relays, PEA Relay Committee Spring Meeting, Matamoras, Pennsylvania; 25-26 May 1995, pp. 1–14.



TABLES

Table 1 Paths And Loops For Solving The Statistical Flowgraph Of Fig. 1

Path	Transmittance
Path 1: 2 → 4	$p_{24}M_{24}(s)$
Path 2: 2 → 3 → 4	$p_{23}p_{34}M_{23}(s)M_{34}(s)$
<i>1st-order loops</i>	
1 → 2 → 1	$p_{12}p_{21}M_{12}(s)M_{21}(s)$
2 → 3 → 2	$p_{23}p_{32}M_{23}(s)M_{32}(s)$
2 → 4 → 5 → 2	$p_{24}p_{45}p_{52}M_{24}(s)M_{45}(s)M_{52}(s)$
2 → 3 → 4 → 5 → 2	$p_{23}p_{34}p_{45}p_{52}M_{23}(s)M_{34}(s)M_{45}(s)M_{52}(s)$
1 → 2 → 4 → 5 → 1	$p_{12}p_{24}p_{45}p_{51}M_{12}(s)M_{24}(s)M_{45}(s)M_{51}(s)$
1 → 2 → 3 → 4 → 5 → 1	$p_{12}p_{23}p_{34}p_{45}p_{51}M_{12}(s)M_{23}(s)M_{34}(s)M_{45}(s)M_{51}(s)$

Table 2 Waiting Time Distributions For The Statistical Flowgraph Of Fig. 1

Transmittance	Distribution	MGF
1 → 2	Inverse Gaussian ($\lambda_{1_{12}}, \lambda_{2_{12}}$)	$M_{12}(s) = \exp \left[\lambda_{1_{12}} \lambda_{2_{12}} - \sqrt{\lambda_{1_{12}}^2 (\lambda_{2_{12}}^2 - 2s)} \right]$
2 → 1	Exponential (μ_{21})	$M_{21}(s) = \left(\frac{\mu_{21}}{\mu_{21} - s} \right)$
2 → 3	Gamma (α_{23}, β_{23})	$M_{23}(s) = \left(\frac{\beta_{23}}{\beta_{23} - s} \right)^{\alpha_{23}}$
2 → 4	Exponential (μ_{24})	$M_{24}(s) = \left(\frac{\mu_{24}}{\mu_{24} - s} \right)$
3 → 2	Exponential (μ_{32})	$M_{32}(s) = \left(\frac{\mu_{32}}{\mu_{32} - s} \right)$
3 → 4	Exponential (μ_{34})	$M_{34}(s) = \left(\frac{\mu_{34}}{\mu_{34} - s} \right)$
4 → 5	Gamma (α_{45}, β_{45})	$M_{45}(s) = \left(\frac{\beta_{45}}{\beta_{45} - s} \right)^{\alpha_{45}}$
5 → 2	Exponential (μ_{52})	$M_{52}(s) = \left(\frac{\mu_{52}}{\mu_{52} - s} \right)$
5 → 1	Exponential (μ_{51})	$M_{51}(s) = \left(\frac{\mu_{51}}{\mu_{51} - s} \right)$

FIGURES

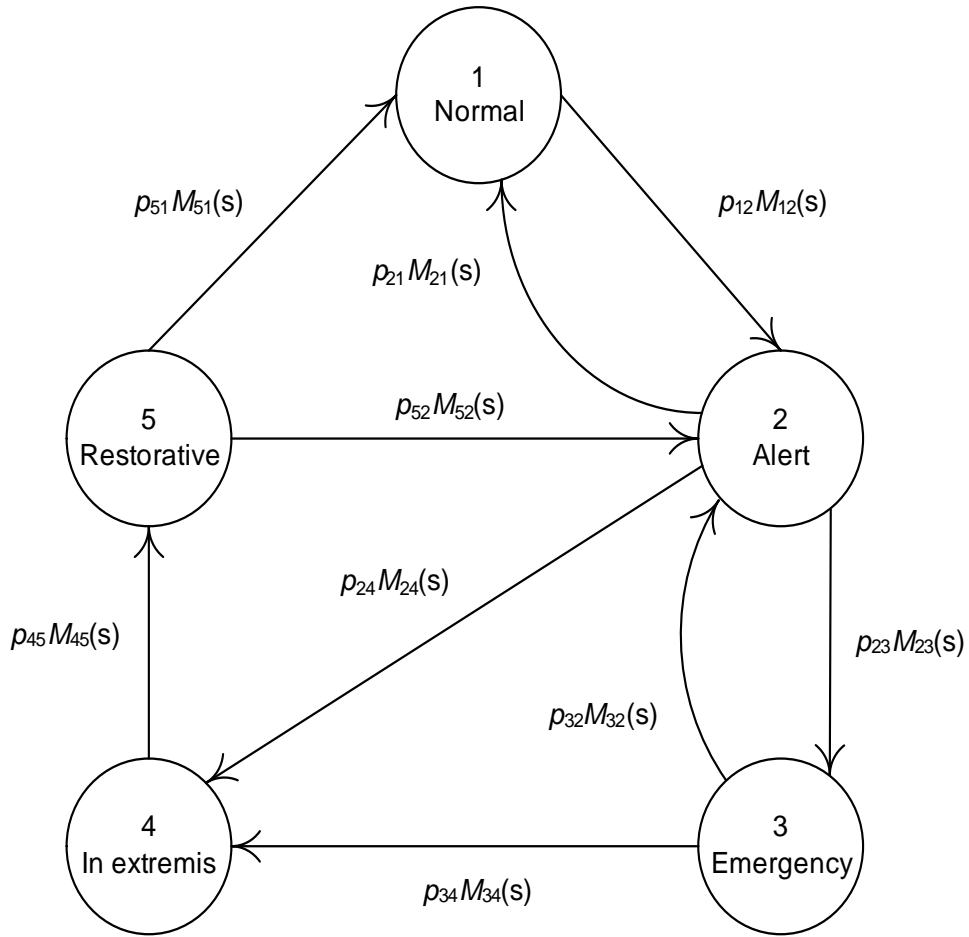


Fig. 1. SFGM For Operating States Of A Power System.

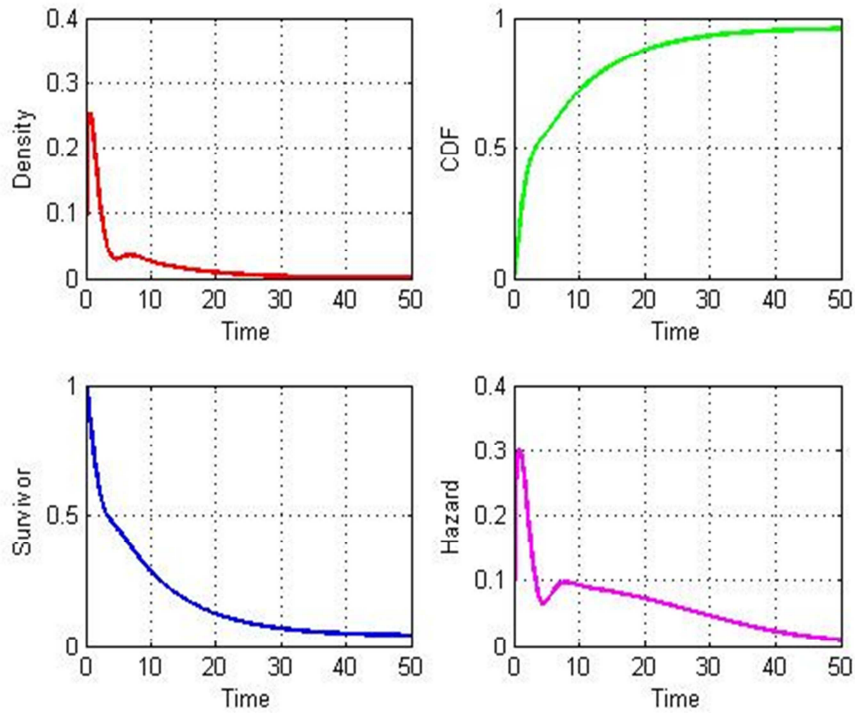


Fig. 2. Density Function, CDF, Survivor Function And Hazard Function For Example 1.

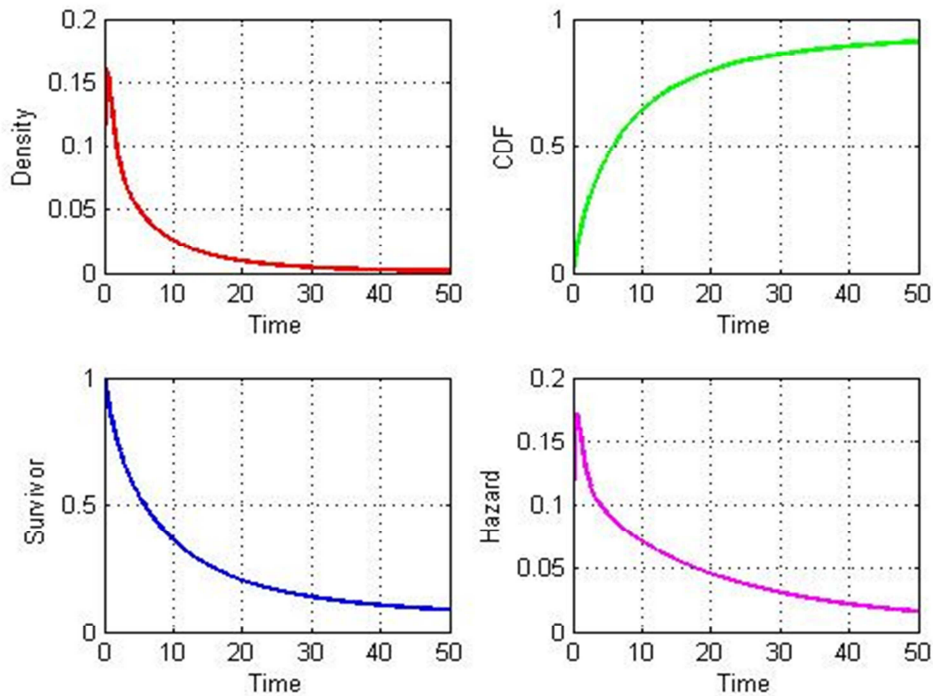


Fig. 3. Bayes Predictive Density Function, CDF, Survivor Function And Hazard Function For Example 1.

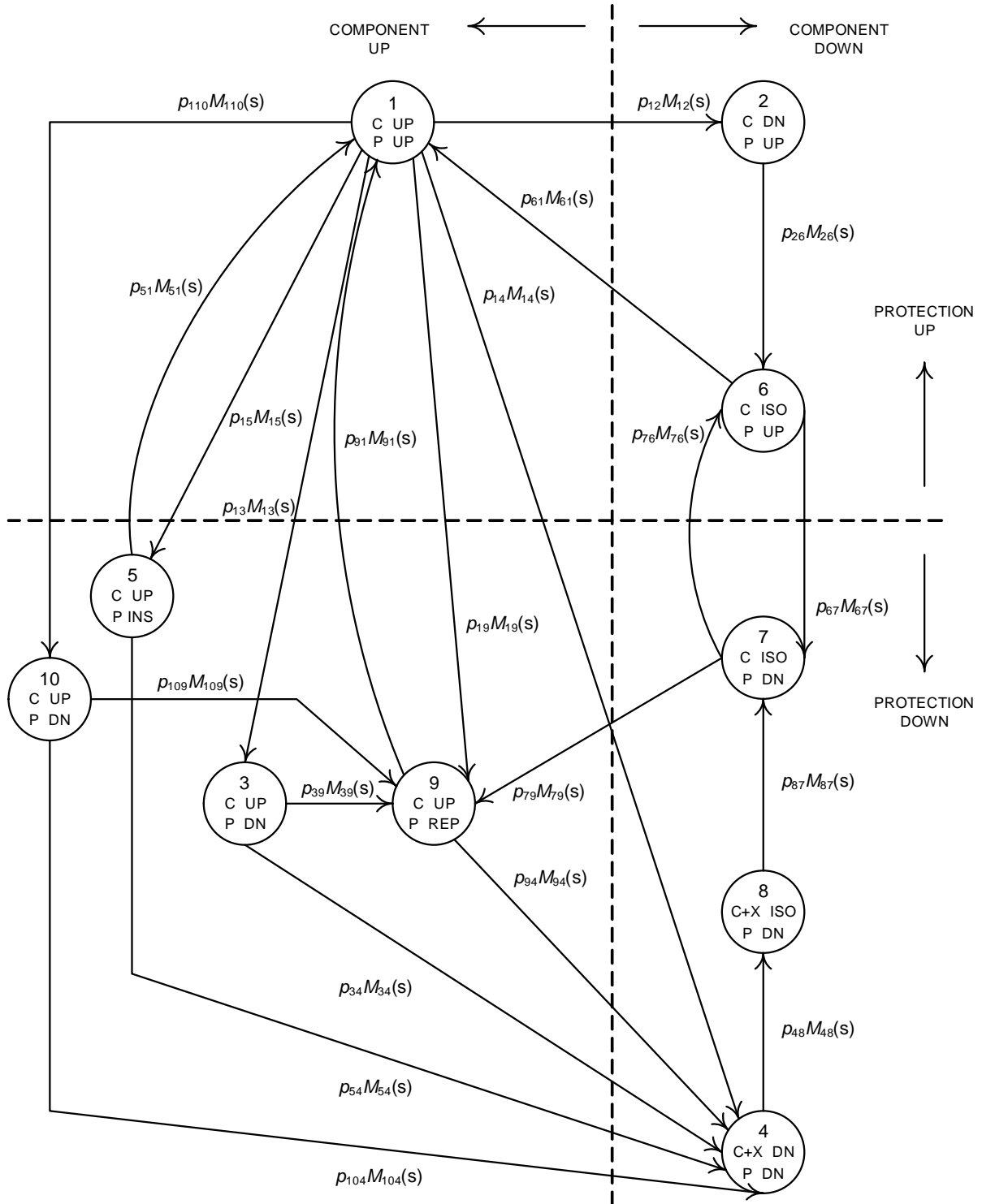


Fig. 4. SFGM For A Protection/Component System.

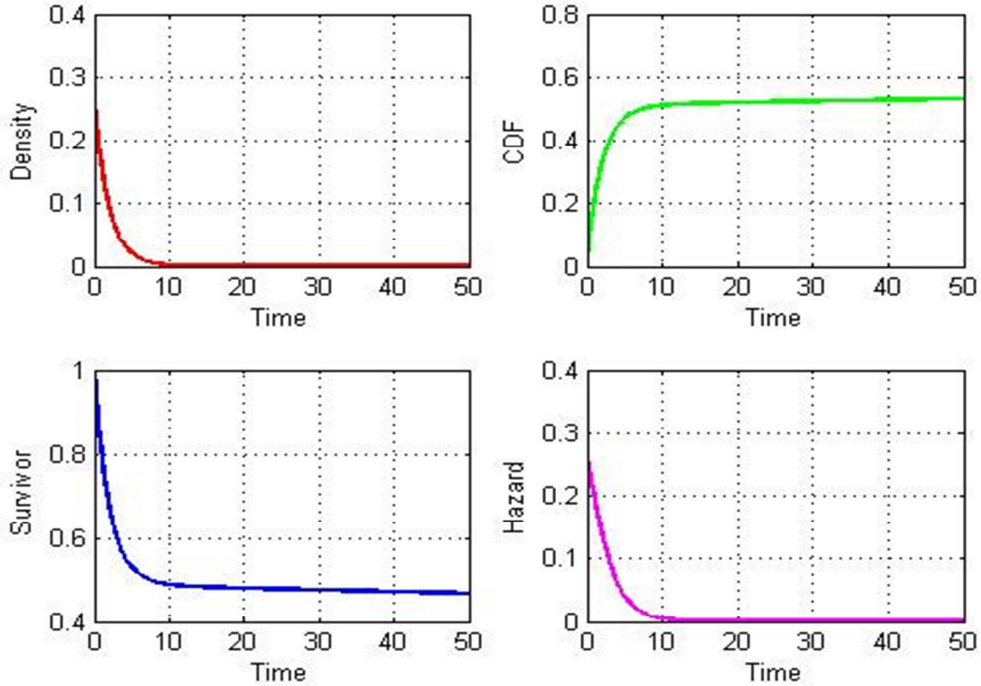


Fig. 5. Density Function, CDF, Survivor Function And Hazard Function For Example 2.

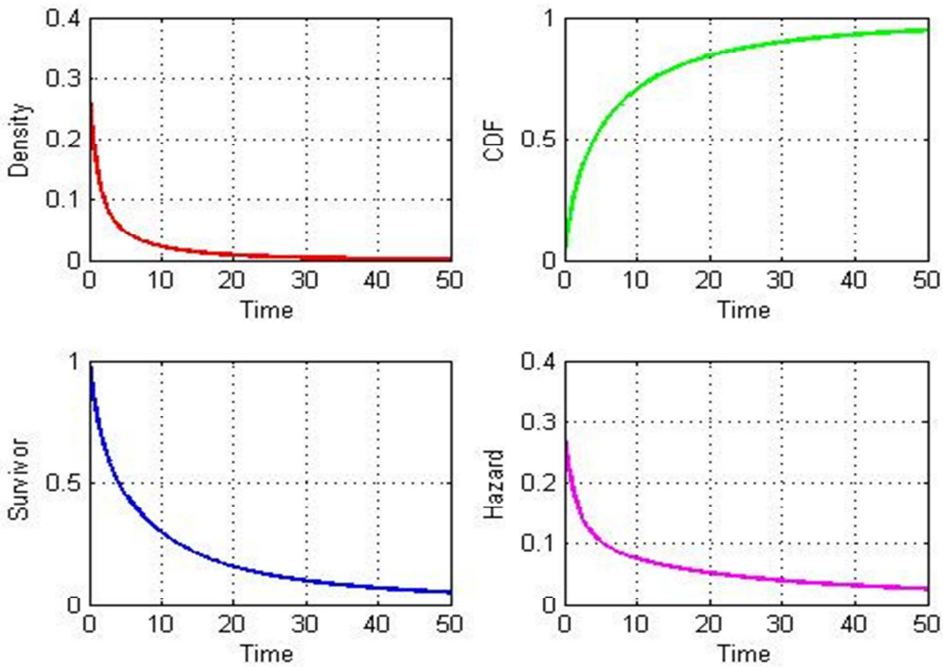


Fig. 6. Bayes Predictive Density Function, CDF, Survivor Function And Hazard Function For Example 2