7-8-2009

Nonparametric estimation of first passage time distributions in flowgraph models

David Collins

David H. Collins

Candidate

Mathematics and Statistics

Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Aparna V. Huzurbazar, Chairperson

David H. Collins

James A. Ellison

Edward D. Graham, Jr.

Nonparametric Estimation of First Passage Time Distributions in Flowgraph Models

by

David H. Collins

B.A., Computational Mathematics, Fordham University, 1983
M.S., Statistics, University of New Mexico, 2005

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Statistics

The University of New Mexico

Albuquerque, New Mexico

May, 2009

Acknowledgements

I would like to thank my entire dissertation committee for helpful and constructive comments on this dissertation. In addition, thanks to Ron Christensen and Jim Ellison for teaching me most of what I know about probability and statistical inference; to Ed Graham for moral support and many useful bits of mathematical and engineering knowledge; to Donna George, Doug Weintraub, and Roxanne Littlefield for helping me to navigate the bureaucracy; and to all my fellow students and the Mahematics and Statistics faculty for a wonderful learning experience. Thanks especially to my advisor, Aparna Huzurbazar, for everything I know about flowgraphs, and for constant mentoring and encouragement since the beginning of my graduate career. Her unselfish engagement with students has been an inspiration. Above all, thanks to my wife, Linn Marks Collins, for supporting and encouraging me through the long process of graduate education.

Nonparametric Estimation of First Passage Time Distributions in Flowgraph Models

by

David H. Collins

ABSTRACT OF DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Statistics

The University of New Mexico

Albuquerque, New Mexico

May, 2009

Nonparametric Estimation of First Passage Time Distributions in Flowgraph Models

by

David H. Collins

B.A., Computational Mathematics, Fordham University, 1983
M.S., Statistics, University of New Mexico, 2005

Ph.D., Statistics, University of New Mexico, 2009

Abstract

Statistical flowgraphs represent multistate semi-Markov processes using integral transforms of transition time distributions between adjacent states; these are combined algebraically and inverted to derive parametric estimates for first passage time distributions between nonadjacent states. This dissertation extends previous work in the field by developing estimation methods for flowgraphs using empirical transforms based on sample data, with no assumption of specific parametric probability models for transition times. We prove strong convergence of empirical flowgraph results to the exact parametric results; develop alternatives for numerical inversion of empirical transforms and compare them in terms of computational complexity, accuracy, and ability to determine error bounds; discuss (with examples) the difficulties of determining confidence bands for distribution estimates obtained in this way; develop confidence intervals for moment-based quantities such as the mean; and show how

methods based on empirical transforms can be modified to accommodate censored data. Several applications of the nonparametric method, based on reliability and survival data, are presented in detail.

Contents

List of Figures

List of Tables

List of Theorems and Lemmas

Preface

To a certain extent this dissertation represents a log of the author's exploration in the fascinating world of flowgraphs, and every entry may not be of interest to every reader. In particular, Section 1.3 is a review of parametric flowgraphs that may be skipped by the reader already familiar with them, and Section 1.4 sketches proofs of parametric flowgraph results that are important, but that the reader may be willing to take on faith. In addition, in some places (e.g., Section 2.2.4, on possible Bayesian approaches to nonparametric flowgraphs) we have described things which are part of a future research agenda rather than completed work.

Flowgraph models offer a particularly attractive combination of conceptual transparency, mathematical elegance, interesting statistics, and practical applications which, it is hoped, the dissertation conveys.

Where it does not conflict with precision, for readability we have avoided abstraction and adopted a style in the vein of applied mathematics and engineering. For example, the stochastic process variable t is treated as "time," which is appropriate for most applications.

Definitions and an index for abbreviations, special functions, and other notation are provided in the Glossary at the end of this document. The symbol ∎ is used to mark the end of a proof.

Chapter 1

Introduction

This chapter defines the problem space addressed in the dissertation and provides necessary background information.

A common situation in reliability analysis is to have data on component failure times (e.g., from unit testing), and to seek a statistical model for predicting the reliability of a system composed of multiple components, perhaps including redundancy and the possibility of repair or replacement of failed components. With the appropriate changes in terminology, similar situations arise in biomedical survival analysis, e.g., substituting states of disease progression for component failures and remission or successful treatment for repair. In contrast to simple two-state models for reliability of nonrepairable components or survival models where disease onset invariably results in fatality, multistate statistical models facilitate more realistic modeling, but raise issues of available theory as well as computational tractability (Hougaard 1999). Section 1.1 presents a simple example of such a multistate model.

Statistical flowgraphs (Huzurbazar 2005a) encompass a body of theory and computational algorithms that has proven successful in modeling complex multistate systems, with applications ranging from reliability and risk analysis (Huzurbazar 2005c; Huzurbazar & Williams 2005) to biomedical survival analysis (Yau & Huzur-

bazar 2002). Prior work in flowgraphs has been largely based on the assumption of parametric probability models for transitions between states, which are incorporated into flowgraph models in the form of integral transforms of their density functions.

In this dissertation we present and analyze a nonparametric approach to using statistical flowgraphs in the analysis of multistate models, which makes no assumptions about the underlying probability models or existence of densities. This approach also allows for semiparametric analysis, with assumed parametric models for some, but not all, of the transitions in a multistate system.

Section 1.2 of this Introduction specifies in more detail the problem domain, and the general problem of estimation (parametric and nonparametric) for multistate semi-Markov models. Sections 1.3 and 1.4 review "classical" (parametric) statistical flowgraphs and their mathematical foundations. Section 1.5 briefly overviews the nonparametric methodology, which uses empirical transforms derived from sample data for solving flowgraphs.

Chapter 2, the core of the dissertation, presents the nonparametric methodology in detail. Chapter 3 illustrates use of the nonparametric methodology through several examples.

1.1 Introductory example

Figure 1.1 shows an example of the type of problem addressed here, a model of a repairable system with redundancy: two units operate in tandem, with the possibility of repair if one fails; system failure (state 3, an absorbing state) occurs if both fail simultaneously, or if the backup unit fails while the primary is being repaired. Assuming the process starts in state 1, interest lies in estimating the probability distribution, or moments of the distribution, for the time of first passage from state 1 to state 3. With obvious differences in terminology, this model also arises in

survival analysis as the reversible illness-death model (Andersen 2002). Transitions are labeled with transition probabilities and integral transforms of the transition time distributions—this is explained in detail in Section 1.3.

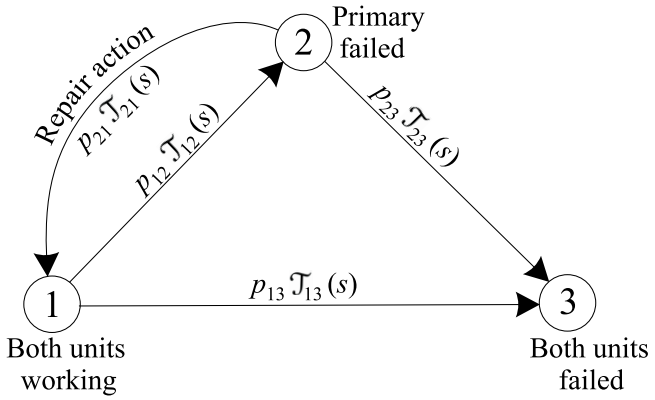


Figure 1.1: Three-state repairable redundant system model

Chapter 3 contains other examples from reliability and survival analysis that illustrate the problem space and show its importance.

1.2 Problem specification and notation

We consider finite-state time-homogeneous semi-Markov processes (Howard 1971b; Çinlar 1975; Ross 1996) represented as graphs (flowgraphs of the processes) whose vertices represent states $1, 2, \ldots, n$ and whose edges represent transitions between states. These are quite general, since a large number of stochastic process models used in applications are specializations of semi-Markov processes: renewal processes, including homogeneous Poisson processes; Markov processes (where transition time distributions are restricted to be exponential), including birth-death processes and Markovian queueing models; and Markov chains (where transition times are identically 1), with the important special cases of discrete random walks and branching

processes. Recurrent-event models, much used in biomedical applications, can often be formulated as semi-Markov models and represented as flowgraphs (Hougaard 1999).

States of a semi-Markov process, considered at transition times, form an embedded Markov chain with transition probabilities p_{ij}. Given a transition from i to j, the holding time in i (transition time from i to j) has distribution function $F_{ij}(t;\theta_{ij})$, where θ_{ij} is a scalar or vector parameter. When dependence on θ_{ij} is understood and we are not interested in the particular parameter value, we omit it from the notation and simply write $F_{ij}(t)$.

The only observable of the process is the state at a given time, so we assume there are no self-transitions (transitions from a state to itself), since these would be unobservable. (However, self-transitions may be introduced by the process of flowgraph reduction; see Section 1.3.) This is without loss of generality since a process with self-transitions can be redefined as one without such transitions, by incorporating the self-transition holding time for each state into the holding time distributions for non-self transitions from that state (Pyke 1961a, Lemma 2).

Define $Q_{ij}(t) = p_{ij}F_{ij}(t)$; the matrix $\mathbf{Q}(t)$ with elements $Q_{ij}(t)$ is called the semi-Markov kernel. $H_i(t) = \sum_{j=1}^{n} Q_{ij}(t)$ is the holding time distribution in i, independent of the state to which a transition is made. Alternatively but equivalently, semi-Markov processes can be viewed as vectors of counting processes $[N_1(\tau) \dots N_n(\tau)]$, where $N_i(\tau)$ is the number of visits to state i at time τ; these are called Markov renewal processes (Pyke 1961a; Ross 1970; Çinlar 1975). We assume that all of our processes are nonexplosive, i.e., that $N_i(\tau)$ is finite for $\tau < \infty$.

To avoid ambiguity, we use the variable t to designate time elapsed since entry to a given state, and τ to designate "calendar time," the time elapsed since the process was started. These times will always be nonnegative.

Where $Z(\tau) \in \{1, 2, \ldots, n\}$ is the state of the process at time $\tau \geq 0$, define the distribution function for first passage time (or hitting time) from i to j as

$$G_{ij}(\tau) = \Pr\{N_j(\tau) > 0 \mid Z(0) = i\},$$

i.e., $G_{ij}(\tau)$ gives the probability that starting in state i, the first visit to state j occurs at or before τ. The corresponding matrix is $\mathbf{G}(\tau)$. For all first passages of interest, we assume that for any $\epsilon > 0$ there exists τ such that $1 - G_{ij}(\tau) < \epsilon$, i.e., that with probability 1, first passages occur in finite time. In a situation (as described in the next paragraph) where a process contains an absorbing state that is irrelevant to the first passage to the state of interest, this is equivalent to redefining G_{ij} as

$$G_{ij}(\tau) = \Pr\{N_j(\tau) > 0 \mid Z(0) = i, \tau < \infty\},$$

the distribution conditional on reaching j in finite time.

We assume there are no *irrelevant states*, states that appear on no path with nonzero probability from i to j, if $i \to j$ is the passage of interest. Figure 1.2 shows an example where state 5 is irrelevant to the $1 \to 4$ passage. Assuming there are no such states is without loss of generality, given the assumption that first passages occur in finite time. Irrelevant states can be eliminated by removing transitions to them and renormalizing probabilities for the remaining transitions; e.g., in Figure 1.2 we would remove the $3 \to 5$ transition and set $p'_{32} = \dfrac{p_{32}}{p_{32} + p_{34}}$, $p'_{34} = \dfrac{p_{34}}{p_{32} + p_{34}}$, where the primed probabilities are normalized. Remaining holding time distributions, being determined by the origin and destination states, do not change.

In the language of Markov chains, we are assuming that our processes have a set of communicating transient states, and a single absorbing state reachable in finite time from any transient state.

The main problem dealt with in this dissertation is the estimation of $G_{ij}(t)$ or its density, given a nonparametric estimate of $\mathbf{Q}(t)$ based on observation of the process.

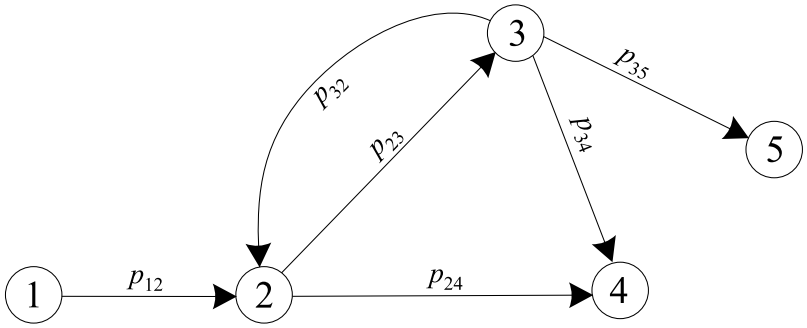


Figure 1.2: Flowgraph of a process with an irrelevant state

1.2.1 Estimation based on observed transition data

The basis for estimating **Q** is sample data on transitions between states. The topological structure of the model (states and possible transitions) may be given or derived from the data. Parametric distribution families for holding times may be given *a priori* (e.g., from the science of the subject domain), inferred from examination of the data, or not used, as in the nonparametric method.

In terms of the information given or assumed about holding times, there is a distinction between having data on transitions between adjacent states only, and having complete event histories. In the former case we may have, say, sample data for the $i \to j$ and $j \to k$ transitions separately, but cannot correlate these into data for the $i \to j \to k$ path. In the latter case the data are given in the form of complete sample paths between states of interest for different subjects or process realizations. (Actually, there is a range of situations between these two cases, where sample paths may be given in various forms of incompleteness; see (Williams & Huzurbazar 2006) for examples. In this dissertation we consider only situations where the given data pertains to transitions between adjacent states.)

When complete event histories are observed for N subjects, a sample of first

passage times $\{\tau_1, \tau_2, \ldots, \tau_N\}$ from i to k is available and the standard nonparametric estimate of G_{ik} is the empirical distribution function (EDF)

$$\hat{G}_{ik}(\tau) = \frac{1}{N} \sum_{m=1}^{N} I_{[0,\tau]}(\tau_m)$$

where I is the indicator function, for which large sample properties such as confidence bands are well known. If a parametric distribution family is assumed for G_{ik}, the method of maximum likelihood provides estimates for the parameter(s) with known large sample properties. These are classical methods which do not require flowgraphs or other techniques specific to stochastic processes.

The more interesting and important case (which we assume hereafter) is where data are available only for adjacent-state transitions. Situations where this arises include:

1. Transition probabilities and holding time distributions can be postulated based on physical models from the application domain, which are in a sense independent of the stochastic model under analysis. An example of this is a model for cumulative seismic damage developed by Gusella (1998) and described in (Collins and Huzurbazar 2008) and in Section 3.3 of this dissertation, where transition probabilities between adjacent states are based on a model for structural damage under shocks, and holding time distributions are given by a separate model for earthquake interarrival times.

2. Data are available from unit testing of components or some equivalent scenario, and in the system context component failures initiate transitions to adjacent states. If more than one transition is possible from a given state, transition probabilities may not be estimatable directly from the data, and holding time distributions may only be estimatable conditionally on the absence of other transitions. However, p_{ij} and F_{ij} can be derived by a competing risks analysis—see Section 1.2.2 below.

3. Transitions between pairs of adjacent states have been studied separately; either different test cases were used for each state pair, or for some other reason data are not available to correlate the adjacent state transitions into complete sample paths. E.g., different failure modes of a complex system have been evaluated in separate tests, individual workstations in a manufacturing process are studied separately, or separate studies have been done on specific stages in the progression of a disease. Examples of this and (2) include risk analyses for large, complex systems that seldom fail, such as nuclear power plants (McCormick 1981; Bedford & Cooke 2001). In the medical field, many studies of HIV-AIDS have focused on transitions between adjacent states in the CDC and WHO classifications; e.g., see (Guihenneuc-Jouyaux *et al.* 2000) and references therein.

Case (1) is inherently parametric, but nonparametric analysis may be used for validation; (2) and (3) may be bases for nonparametric analysis, or used as starting points for parametric analysis by hypothesizing a model using a histogram or other density estimate, which is then fitted using maximum likelihood. Here our main interest is in nonparametric analysis.

If complete event histories are observed for an interval $[0, \tau]$, with $\hat{N}_{ij}(\tau)$ the observed number of $i \to j$ transitions and $\hat{N}_i(\tau)$ the observed number of visits to i, the semi-Markov kernel is estimated as (Moore & Pyke 1968)

$$\hat{p}_{ij} = \frac{\hat{N}_{ij}(\tau)}{\hat{N}_i(\tau)}, \quad \hat{F}_{ij}(t) = \frac{1}{\hat{N}_{ij}(\tau)} \sum_{k=1}^{\hat{N}_{ij}(\tau)} I_{[0,t]}(k\text{th } i \to j \text{ transition time}),$$

$$\hat{Q}_{ij}(t) = \hat{p}_{ij}\hat{F}_{ij}(t). \tag{1.1}$$

If the observed holding times for the transition $i \to j$ are $t_1, \ldots, t_{\hat{N}_{ij}(\tau)}$, then the

empirical mass function corresponding to $\hat{F}_{ij}$ is

$$\hat{f}_{ij}(t) = \frac{1}{\hat{N}_{ij}(\tau)} I_{\{t_1,\ldots,t_{\hat{N}_{ij}(\tau)}\}}(t).$$

Under the assumption that we do not observe complete event histories, $\hat{N}_{ij}(\tau)$ is the number of subjects or test units, but $\hat{N}_i(\tau)$ is not observed. In some cases of incomplete observation $\hat{N}_i(\tau)$ is unnecessary, since what is really needed is the ratio $\hat{N}_{ij}(\tau)/\hat{N}_i(\tau) = \hat{p}_{ij}$; e.g., where transitions from i occur only to a single state j, we know that $p_{ij} = 1$. If there are parallel transitions from state i to $j, k, \ldots$, we must infer $\hat{p}_{ij}$ and $\hat{F}_{ij}$ from a competing risks analysis as described below.

Moore and Pyke's estimator (1.1) assumes the data are uncensored; analogously Gill (1980a) uses the Kaplan-Meier estimator of the survival function (Kaplan & Meier 1958) to develop an estimator for $\hat{Q}_{ij}$ when the sample data are randomly right-censored.

1.2.2 Competing risks analysis

In cases (2) and (3) above, where state transitions are studied individually, It is not possible in general to calculate all the quantities in (1.1) directly. For example, in the repairable system portrayed in Figure 1.1, unit testing would typically provide data on the $2 \to 3$ failure transition but not on the $2 \to 1$ repair transition. A separate study might provide sample data from the distribution of repair times.

This is a competing risks scenario (Crowder 2001), illustrated in the general case by Figure 1.3. Define $T_{i1}, \ldots, T_{ij}, \ldots$ as the *latent holding times*, T_{ij} being the $i \to j$ holding time when all other transitions have been removed. What we actually observe is T_{ij} conditional on $T_{ij} = \min_k(T_{ik})$ (called the "crude" lifetimes in actuarial terminology). The T_{ij} have a joint distribution function $F_i(\ldots, t_{ij}, \ldots)$ whose marginals $F_i^j(t_{ij})$ are the latent (or "net") holding time distributions.

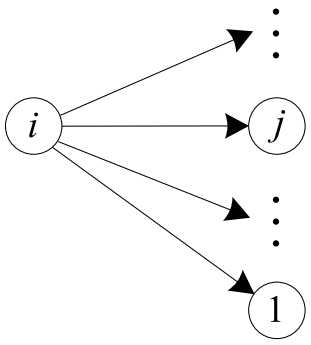


Figure 1.3: Transitions in parallel as competing risks

In the example of state 2 in Figure 1.1, from the testing data we can produce estimates of $F_2^3(t_{23})$ and $F_2^1(t_{21})$; what we need are estimates of

$$p_{23} = \mathrm{P}\{2 \to 3 \mid 2\} = \mathrm{P}\{T_{23} < T_{21}\} \text{ and}$$
$$F_{23}(t) = \mathrm{P}\{T_{23} \le t \mid 2 \to 3\} = \mathrm{P}\{T_{23} \le t \mid T_{23} < T_{21}\},$$

and likewise for p_{21} and F_{21}.

In the parametric case, suppose $f_2(t_{21}, t_{23})$ is the joint density of T_{21} and T_{23}. Then p_{23} is calculated by integrating $f_2(\cdot, \cdot)$ over the region where $T_{23} < T_{21}$:

$$p_{23} = \int_0^\infty \int_0^{t_{21}} f_2(t_{21}, t_{23}) dt_{23} dt_{21}.$$

The distribution function of T_{23} is obtained by integrating over the region where $T_{23} < T_{21}$ and $T_{23} \le t$, and normalizing:

$$F_{23}(t) = \frac{1}{p_{23}} \int_0^t \int_0^{t_{21}} f_2(t_{21}, t_{23}) dt_{23} dt_{21}.$$

F_{21} and p_{21} are calculated similarly, and the same idea extends to states with more than two outbound transitions. See Section 2.2 of (Huzurbazar 2005a) for further examples of parametric competing risks analysis.

Nonparametric estimators for p_{ij} and F_{ij} can be derived by bootstrapping: resample the samples of $T_{i1}, \ldots, T_{ij}, \ldots$ to get bootstrap samples $\{t_{i1}^*, \ldots, t_{ij}^*, \ldots\}$, select

those t^*_{ij} which are minima of their respective bootstrap samples (say n of them), then estimate $F_{ij}(t)$ as

$$\hat{F}_{ij}(t) = \frac{1}{n} \sum_{t^*_{ij}} I_{[0,t]}(t^*_{ij}),$$

and $\hat{p}_{ij}$ as the proportion of samples in which t^*_{ij} is the minimum. This assumes there are no samples with tied observations. A simple way of breaking ties is to jitter the data by adding iid perturbation terms, distributed uniformly over $(-\epsilon, \epsilon)$ for small ϵ, to the observations (Owen 2001, Section 3.7). An alternative is the smoothed bootstrap (Davison & Hinkley 1997, Section 3.4), which resamples from smoothed approximations of the empirical distribution functions based on the samples.

For censored data, the bootstrapping procedure above can be carried out using the Kaplan-Meier estimates of the EDFs. Alternative methods are described in (Davison & Hinkley 1997, Section 3.5).

1.3 Parametric statistical flowgraph models

This Section and the next provide a review of flowgraphs as they have been used to model stochastic processes using parametric distribution families. For more information see (Huzurbazar 2005a) and references therein.

Statistical flowgraphs are graphical representations of the states and transitions of semi-Markov processes characterized by interstate transition probabilities p_{ij} and holding time distribution functions $F_{ij}(t; \theta_{ij})$, assumed to be members of parametric families such as gamma or Weibull. We usually omit the parameter and write $F_{ij}(t)$.

The holding time distribution between states i and j is represented by its transform, typically either the Laplace transform (LT) $\mathcal{L}_{ij}(s) = \int_0^\infty e^{-st} dF_{ij}(t)$ or the moment generating function (MGF) $\mathcal{M}_{ij}(s) = \int_0^\infty e^{st} dF_{ij}(t)$. The Fourier transform (characteristic function or CF) $\varphi_{ij}(s) = \int_0^\infty e^{ist} dF_{ij}(t)$ is also useful since it always

exists, whereas the LT and MGF fail to converge for some probability densities such as the lognormal. (The lower limit of 0 for the CF reflects the assumption that all the random variables we study are non-negative. Mathematicians usually write the Fourier transform with a negative sign on the exponent, and may also have a multiplier of $1/2\pi$ or $1/\sqrt{2\pi}$; these differences affect the relationship between the formulas for the transform and its inverse, and will not concern us.)

We use $\mathcal{T}_{ij}(s)$ to designate a general integral transform (as in the labeling of Figure 1.1). Thus a flowgraph consists of a vertex set $\{i\}$ representing states, and an edge set of pairs $\{(p_{ij}, \mathcal{T}_{ij})\}$ representing transitions. In this dissertation, unless stated otherwise, $\mathcal{T}_{ij}(s)$ will be either the LT or MGF.

In the literature on flowgraphs it is standard to assume that holding time distributions have continuous probability densities (pdfs), i.e., that the distributions are absolutely continuous with respect to Lebesgue measure. Thus if the pdf is $f_{ij}(t)$ and the transform has a continuous kernel ψ, $\mathcal{T}_{ij}(s) = \int_0^\infty \psi(t,s) f_{ij}(t) dt$, where the integral is the usual Riemann integral. Solving flowgraphs involves computations on the transforms, which motivates this simplifying assumption, particularly in the engineering literature. We will stick with the assumption for the moment; in Section 1.4.5 we show that it involves no real loss of generality, since all the computations work the same way when we base the transforms on (cumulative) distribution functions (CDFs) and use more general types of integral.

Flowgraph models first appear in the literature of electrical engineering as "signal flow graphs" (Mason 1953) in which vertices represent points at which a signal can be measured, and each edge represents an active element with *transmittance* characterized by a transfer function, the Laplace transform of the element's response to an input signal.

1.3.1 Transforms, transmittances and Mason's rule

In a statistical flowgraph the transmittance between adjacent states i and j is the product of the transition probability p_{ij} and the transform $\mathcal{T}_{ij}(s)$ of the holding time distribution; in the notation of Section 1.2, $p_{ij}\mathcal{T}_{ij}(s)$ is the transform of $Q_{ij}(t) = p_{ij}F_{ij}(t)$.

Mathematically, this transmittance is a transform; if $\mathcal{T}(s) = \int_0^\infty \psi(t,s)dF(t)$ is a transform of F, then for $\lambda \in \mathbb{R}$, $\lambda\mathcal{T}(s) = \int_0^\infty \psi(t,s)d[\lambda F](t)$ is the same transform of λF. Transmittances between non-adjacent states are also transforms; the general equivalence of transmittances and transforms will be shown later (see Section 2.1.2). Where it is desirable to make a distinction between transforms and transmittances, usually because we wish to differentiate between simple transforms of distributions and transmittances compounded from multiple transforms, we will symbolize transmittances as $\mathfrak{T}(s)$.

All transforms $\mathcal{T}(s)$ used here within the scope of a proof or computation will be the same type (normally either LTs or MGFs). For nonparametric flowgraph methods, as shown in Chapter 2 the empirical versions of the LT and MGF will always converge, so the advantage mentioned above for the CF is only important when using parametric flowgraph methods. The disadvantage of the CF is the fact that it is in general complex-valued, which makes proofs more difficult; see, e.g., (Lukacs 1960).

Supposing we have transition times with CDFs $F, G, \ldots$, for the computations used in solving for first passage distributions we require the following properties of the transforms:

Linearity: $\mathcal{T}_{\alpha F+\beta G}(s) = \alpha\mathcal{T}_F(s) + \beta\mathcal{T}_G(s)$, where $\mathcal{T}_F$ is the transform of F and α, β are scalars.

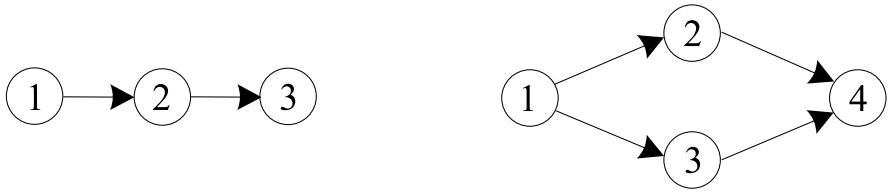


Figure 1.4: Basic elements for flowgraph analysis: convolution, mixture

Figure 1.5: Basic elements for flowgraph analysis: loop, reduced loop

Convolution: $\mathfrak{T}_{F\star G}(s) = \mathfrak{T}_F(s)\mathfrak{T}_G(s)$, where $F \star G(t) = \int G(t-u)dF(u)$.

Uniqueness: If $\mathfrak{T}(s)$ converges in a neighborhood $-\delta < s < \delta$, $\delta > 0$, it defines a probability distribution uniquely up to sets of measure zero. Thus if $\mathfrak{T}_1(s) = \mathfrak{T}_2(s)$ for s in the interval of convergence, they are transforms of the same distribution.

These are satisfied by the LT, MGF, and CF, and by the empirical transforms defined in Chapter 2. In fact, it can be shown (Lukacs 1960, Theorem 4.4.1) that for any transform with a kernel $\psi(t,s)$ satisfying these properties with $s,t \in \mathbb{R}$, $\psi(t,s) = e^{tA(s)}$, where $A(s)$ is dense in $(0,\infty)$ and real-valued if $\psi(t,s)$ is real-valued, imaginary-valued if $\psi(t,s)$ is complex-valued. Thus it is fair to say that the LT, MGF, and CF exhaust the possibilities for the general form of transform kernels useful for flowgraph modeling.

A flowgraph is solved by computing transmittances between arbitrary pairs of states, then inverting the transforms to recover densities or distribution functions.

Typically interest lies in determining the distribution of the first passage time between an initial state and an absorbing final state. (These are called the source and sink nodes in the electrical engineering literature.) The method consists of recursive application of three rules that reduce transmittances over complex paths constructed from the building blocks shown in Figures 1.4 and 1.5, to a single equivalent transmittance:

1. The transmittance of edges in series is the product (convolution) of the edge transmittances. This is the standard convolution theorem for sums of random variables (Casella & Berger 2002, Theorem 5.2.9). E.g., on the left in Figure 1.4, since $p_{12} = p_{23} = 1$, the edge transmittances are just the transforms and

   $$\mathfrak{T}_{13}(s) = \mathfrak{T}_{12}(s)\mathfrak{T}_{23}(s).$$

2. The transmittance of edges in parallel is the sum (mixture) of the edge transmittances. This follows from the linearity of the transform. E.g., on the right in Figure 1.4,

   $$\mathfrak{T}_{14}(s) = p_{12}\mathfrak{T}_{12}(s)\mathfrak{T}_{24}(s) + p_{13}\mathfrak{T}_{13}(s)\mathfrak{T}_{34}(s).$$

3. Rule 1 reduces the graph on the left in figure 1.5 to the one on the right, where the loop transmittance $\mathfrak{T}_L(s) = p_{12}\mathfrak{T}_{12}(s)\mathfrak{T}_{21}(s)$. (As mentioned in Section 1.2, this reduction introduces a self-transition which did not exist in the original flowgraph.) Now let $\mathfrak{T}(s)$ be the $1 \to 3$ first passage transmittance (versus the one-step transmittance $\mathfrak{T}_{13}(s)$). By the Markov property, $\mathfrak{T}(s)$ is unchanged by any number of traversals of the loop, so by this and Rules 1 and 2,

   $$\begin{aligned} \mathfrak{T}(s) &= \mathfrak{T}_L(s)\mathfrak{T}(s) + p_{13}\mathfrak{T}_{13}(s) \\ &= \frac{p_{13}\mathfrak{T}_{13}(s)}{1 - \mathfrak{T}_L(s)}. \end{aligned}$$

Alternatively, this can be seen by enumerating and summing all possible parallel transition paths from 1 to 3:

$$\begin{aligned}\mathfrak{T}(s) &= p_{13}\mathcal{T}_{13}(s) + \mathfrak{T}_L(s)p_{13}\mathcal{T}_{13}(s) + \mathfrak{T}_L(s)^2 p_{13}\mathcal{T}_{13}(s) + \mathfrak{T}_L(s)^3 p_{13}\mathcal{T}_{13}(s) + \dots \\ &= \frac{p_{13}\mathcal{T}_{13}(s)}{1-\mathfrak{T}_L(s)}.\end{aligned}$$

Summing this as a convergent geometric series depends on the fact that for the transform parameter s in some open neighborhood of 0, $0 < |\,\mathfrak{T}_L(s)| < 1$. This is true because $p_{12} < 1$ (otherwise the loop would never be exited), $\mathcal{T}_{12}(0) = \mathcal{T}_{21}(0) = 1$, and both transforms are continuous functions of s.

The reduced transmittance is the transform of the time distribution for first passage between the states of interest. As an illustration, consider the model in Figure 1.1 and let $\mathfrak{T}(s)$ be the transmittance for first passage from state 1 to state 3, independent of the path. Using the rules above,

$$\mathfrak{T}(s) = [p_{12}\mathcal{T}_{12}(s)p_{21}\mathcal{T}_{21}(s)]\mathfrak{T}(s) + p_{12}\mathcal{T}_{12}(s)p_{23}\mathcal{T}_{23}(s) + p_{13}\mathcal{T}_{13}(s)$$

Solving for $\mathfrak{T}(s)$ yields

$$\mathfrak{T}(s) = \frac{p_{12}\mathcal{T}_{12}(s)p_{23}\mathcal{T}_{23}(s) + p_{13}\mathcal{T}_{13}(s)}{1 - p_{12}\mathcal{T}_{12}(s)p_{21}\mathcal{T}_{21}(s)}. \tag{1.2}$$

From these rules a formal algorithm known as *Mason's rule* can be derived:

$$\mathfrak{T}(s) = \frac{\sum_k \left\{\mathfrak{T}_k(s)\left[1 + \sum_m (-1)^m L_m^k(s)\right]\right\}}{1 + \sum_m (-1)^m L_m(s)}. \tag{1.3}$$

In this formula $\mathfrak{T}(s)$ is the transmittance between the states of interest, say i and j. $\mathfrak{T}_k(s)$ is the transmittance of the kth direct path (path with no loops) from i to j. $L_m^k(s)$ in the numerator is the sum of transmittances of all mth-order loops with no states in common with the kth direct path. (A first-order loop is a closed path

with the same start and end states that does not pass through any state more than once. An mth-order loop consists of m first-order loops with no points in common; the transmittance of an mth-order loop is the product of the transmittances of the first-order loops.) $L_m(s)$ in the denominator is the sum of transmittance over all mth-order loops.

If the flowgraph has loops, $\mathfrak{T}(s)$ has one or more singularities (simple poles) at values of s where the denominator vanishes. As mentioned in (3) above, from continuity of the transforms and the fact that loop probabilities must be strictly less than 1, it follows that there are no singularities in some open neighborhood of 0.

Based on what was said at the beginning of this section, $\mathfrak{T}(s)$ is in fact a transform, the transform of the first passage distribution from i to j. This will be shown more rigorously in Sections 1.4 and 2.1.2.

There are various derivations of the rather cryptic formula (1.3); the most straightforward, an application of Cramer's rule to a certain set of linear equations, is sketched in the next section. Mason's rule provides a solution for first passage transforms in arbitrarily complex flowgraphs. For examples of its use with statistical flowgraphs, see (Huzurbazar 2005a).

The flowgraph reduction rules can be viewed more generally as operations on directed graphs with edge weights. Thus it is not surprising to find essentially the same rules used in diverse fields such as finite automata theory (Csenki 2008) and complexity analysis of computer programs (McCabe 1976). We will not pursue the graph-theoretic aspects of these rules except in an informal way via diagrams.

The steps in parametric flowgraph modeling can be summarized as follows:

1. Determine the topology of the directed state graph.
2. Posit a parametric family for each holding time distribution.

3. Estimate parameters of the probability distributions from sample data using maximum likelihood.
4. Estimate transition probabilities based on data or competing risks analysis.
5. Compute a closed-form expression for the transform of each holding time pdf.
6. Use Mason's rule to compute transforms for transitions of interest.
7. Invert these transforms to obtain pdfs. In some cases this can be done analytically; in general, numerical methods must be used.

If holding time data are subject to censoring, appropriate modifications are made for maximum likelihood estimation (Klein and Moeschberger 2003, Chapter 3). Censored data histograms (Huzurbazar 2005b) can be used as an aid in determining parametric models for distributions. For further details on the steps described above, see (Huzurbazar 2005a).

1.4 A rigorous basis for statistical flowgraphs

Mason's (1953, 1956) derivation of his eponymous rule has been characterized as "the most advanced form of graphic algebra" (Robichaud *et al.* 1962, p. viii). Though offering a certain conceptual transparency, Mason's derivation is lengthy, less than rigorous by modern standards, and uses an engineering vocabulary that is unfamiliar to most mathematicians and statisticians. The use of flowgraph methods for analysis of stochastic processes is based on Mason's and other earlier work in the theory of linear time-invariant systems, originating in electrical and control engineering; e.g., see (Howard 1971b) and references therein. The same basis has been used to justify "transform methods" generally for analysis of stochastic processes; e.g., see (Kleinrock 1975).

Many modern texts show a certain distaste for transform methods, called by one author "the oft-lamented *Laplacian curtain*, which covers the solution and hides the structural properties of many interesting stochastic models" (Neuts 1981, p. x). Such aversion may be partly accounted for by the perceived difficulty of transform inversion (Davies & Martin 1979). This however, is no longer a valid objection, since fast and accurate algorithms for numerical inversion of transforms of distribution functions now exist (Abate & Whitt 1992; Strawderman 2004). In addition, we have found that the graphical approach which is enabled by using transforms often illuminates the structure of a stochastic process in ways that purely matrix-based approaches do not.

In this section we sketch a rigorous basis for statistical flowgraph analysis based on the large body of flowgraph literature, both in statistics and engineering. To fill in the details of the sketch, the reader is referred to the literature cited at the end of the section.

The key results are:

1. Determining the transform for a first passage time is equivalent to solving a system of linear equations.
2. The solution of these equations by Cramer's rule is equivalent to Mason's rule.
3. The solution is equivalent to Pyke's (1961b) solution based on Markov renewal theory.

1.4.1 Flowgraphs as systems of linear equations

We consider the simple problem of finding the first-passage transmittance from state i to j; later we will show how this can be generalized to find the first passage transmittance between two disjoint sets of states. The setup is as follows:

1. Renumber the states so that i becomes 1, j becomes n, and the remainder are arbitrary; this simplifies the notation.

2. Remove all outgoing transitions (if any) from state n.

3. Add a fictitious state 0 which transitions to state 1 in zero time with probability 1. The $0 \to 1$ transmittance using the LT is $\int e^{-st}\delta(t)dt = 1$, where $\delta(t)dt$ is Dirac measure. The transmittance is also 1 using the CF or MGF. This captures the fact that the process starts in state 1 at $\tau = 0$.

With this setup, let x_k be the transmittance for passage from $0 \to k$ (not necessarily the first passage), i.e., the transmittance for some feasible sequence of states $0 \to \ldots \to \ldots \to k$. (This is an abuse of notation, used for consistency with the usual way of writing linear equations and with most of the flowgraph literature; we should say $x_k(s)$, since x_k is a transform expression parameterized by s.) The transmittance to k given that m was the previous state is $\mathfrak{T}_{mk}x_m$, so allowing for the transition from state 0,

$$x_1 = 1 + \mathfrak{T}_{11}x_1 + \mathfrak{T}_{21}x_2 + \ldots + \mathfrak{T}_{n1}x_n$$

and for $k \neq 1$,

$$x_k = \quad \mathfrak{T}_{1k}x_1 + \mathfrak{T}_{2k}x_2 + \ldots + \mathfrak{T}_{nk}x_n.$$

Rearranging,

$$\begin{aligned}
(1-\mathfrak{T}_{11})x_1 - \mathfrak{T}_{21}x_2 - \ldots - \mathfrak{T}_{n1}x_n &= 1 \\
-\mathfrak{T}_{11}x_1 + (1-\mathfrak{T}_{22})x_2 - \ldots - \mathfrak{T}_{n2}x_n &= 0 \\
&\vdots \\
-\mathfrak{T}_{1k}x_1 - \mathfrak{T}_{2k}x_2 - \ldots + (1-\mathfrak{T}_{kk})x_k - \ldots - \mathfrak{T}_{nk}x_n &= 0 \\
&\vdots \\
-\mathfrak{T}_{1n}x_1 - \mathfrak{T}_{2n}x_2 - \ldots + (1-\mathfrak{T}_{nn})x_n &= 0.
\end{aligned} \tag{1.4}$$

Note that all the $\mathfrak{T}_{ni}$ are actually zero because transitions out of state n were removed.

$$\text{Now let } \boldsymbol{\mathfrak{T}}(s) = \begin{bmatrix}
0 & \mathfrak{T}_{12}(s) & \ldots & \mathfrak{T}_{1,n-1}(s) & \mathfrak{T}_{1n}(s) \\
\mathfrak{T}_{21}(s) & 0 & \ldots & \mathfrak{T}_{2,n-1}(s) & \mathfrak{T}_{2n}(s) \\
\vdots & & & & \vdots \\
\mathfrak{T}_{n-1,1}(s) & \mathfrak{T}_{n-1,2}(s) & \ldots & 0 & \mathfrak{T}_{n-1,n}(s) \\
0 & 0 & 0 & 0 & 0
\end{bmatrix},$$

the matrix of one-step transmittances for the flowgraph. The diagonal elements are zero since by assumption there are no self-transitions. The last row is zero because transitions out of state n were removed. With this definition, the equations (1.4) can be expressed as

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{1.5}$$

where

$$\mathbf{A} = \begin{bmatrix}
1 & -\mathfrak{T}_{21}(s) & \ldots & -\mathfrak{T}_{n-1,1}(s) & 0 \\
-\mathfrak{T}_{12}(s) & 1 & \ldots & -\mathfrak{T}_{n-1,2}(s) & 0 \\
\vdots & & & \vdots & \\
-\mathfrak{T}_{1n}(s) & -\mathfrak{T}_{2n}(s) & \ldots & -\mathfrak{T}_{n-1,n}(s) & 1
\end{bmatrix}, \ \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Note that $\mathbf{A} = \left[\mathbf{I} - \mathfrak{T}(s)^{\mathrm{T}}\right]$.

Since all outgoing transitions were removed from state n, the only possible transition into n is the first, thus the solution of the system for x_n yields the $1 \to n$ first passage transmittance.

This approach can be generalized to solve for first passage times from each of a set of transient states $U = \{i_1, i_2, \ldots\}$ (renumbered as $\{1, 2, \ldots\}$) by replacing the vector $\mathbf{b}$ with $[\pi_1, \pi_2, ..., 0]^{\mathrm{T}}$, where π_i is the probability of starting in the ith state and $\sum_{i=1}^{n} \pi_i = 1$. It can also be generalized to a process with a set of absorbing states $D = \{j_1, j_2, \ldots\}$ (renumbered as $\{\ldots, n-1, n\}$). This is done by removing the states D from the flowgraph along with their outgoing transitions, and adding a new state δ which becomes the destination for all transitions into states in D. I.e., all the states in D are aggregated into the single absorbing state δ, so solving the linear equations for x_δ gives the first passage transmittance into any state in D. (See (Limnios & Oprişan 2001, Section 5.4) for a similar approach based on Markov renewal processes.)

1.4.2 Existence of solutions

Existence of a unique solution depends on invertibility of $\mathbf{A}$ or equivalently, $\det \mathbf{A} \neq 0$. Since $\det \mathbf{A} = \det \mathbf{A}^{\mathrm{T}}$, it suffices to show that $\det\left[\mathbf{I} - \mathfrak{T}(s)\right] \neq 0$. Expanding the determinant about the last row,

$$\det\left[\mathbf{I} - \mathfrak{T}(s)\right] = (-1)^{n+n} \det\left[\mathbf{I} - \mathfrak{T}(s)\right]_{nn} = \det\left[\mathbf{I} - \mathfrak{T}(s)\right]_{nn}$$

where $\left[\mathbf{I} - \mathfrak{T}(s)\right]_{nn}$ is $\mathbf{I} - \mathfrak{T}(s)$ with the nth row and nth column removed.

Now note that $\left[\mathfrak{T}(s)\right]_{nn}$ is the transmittance matrix for states $1, \ldots, n-1$. Since the 2-step transmittance from i to j is $\sum_{k=1}^{n-1} \mathfrak{T}_{ik}(s)\mathfrak{T}_{kj}(s)$, clearly the matrix of 2-step transmittances is $\left[\mathfrak{T}(s)\right]_{nn}^2$, and by induction the m-step transmittance matrix

is $[\mathfrak{T}(s)]^m_{nn}$. By assumption, states $1, \ldots, n-1$ are transient, so $[\mathfrak{T}(s)]^m \to 0$ elementwise as $m \to \infty$ (Çinlar 1975, Theorem 5.3.2). Then by a standard result in matrix analysis (Kemeny & Snell 1960, Theorem 1.11.1), $[\mathbf{I} - \mathfrak{T}(s)]_{nn}$ is invertible and $[\mathbf{I} - \mathfrak{T}(s)]^{-1}_{nn} = \sum_{m=0}^{\infty} [\mathfrak{T}(s)]^m_{nn}$. So $\det [\mathfrak{T}(s)]_{nn} = \det \mathbf{A} \neq 0$ and a solution of equation (1.5) exists.

1.4.3 Equivalence to Mason's rule

Given the existence of a solution, we have by Cramer's rule

$$x_n = \frac{\det [\mathbf{A}_1 \ldots \mathbf{A}_{n-1}\, \mathbf{b}]}{\det \mathbf{A}} \tag{1.6}$$

where the numerator is the determinant of $\mathbf{A}$ with the nth column replaced by $\mathbf{b}$. By definition,

$$\det \mathbf{A} = \sum_{\sigma} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \ldots a_{n\sigma(n)} \tag{1.7}$$

where the sum is over the $n!$ permutations σ of $1, \ldots, n$ and $\operatorname{sgn}(\sigma)$ is the sign of the permutation, positive for an even number of transpositions of the second subscripts, negative for an odd number.

Considering the denominator first, we transpose to simplify the notation, so the ijth element of $\mathbf{A}^{\mathrm{T}}$ is 1 if $i = j$ and $-\mathfrak{T}_{ij}$ (possibly 0) if not:

$$\det \mathbf{A}^{\mathrm{T}} = \det [\mathbf{I} - \mathfrak{T}(s)] = \det \begin{bmatrix} 1 & -\mathfrak{T}_{12}(s) & \ldots & -\mathfrak{T}_{1,n}(s) \\ -\mathfrak{T}_{21}(s) & 1 & \ldots & -\mathfrak{T}_{2,n}(s) \\ \vdots & & & \vdots \\ -\mathfrak{T}_{n-1,1}(s) & -\mathfrak{T}_{n-1,2}(s) & \ldots & -\mathfrak{T}_{n-1,n}(s) \\ 0 & 0 & \ldots & 1 \end{bmatrix}. \tag{1.8}$$

There is one term in the determinant consisting of all the diagonal factors, with a value of +1. In the remaining terms we may ignore the diagonal factors, since their contribution of +1 does not change the value of the term. Every integer in $\{1, \ldots, n\}$ occurs once in some term as first subscript and once as second subscript, so given a factor $a_{k_1\sigma(k_1)}$ with $k_1 \neq \sigma(k_1)$ there must be another factor $a_{\sigma(k_1)k_2}$, where $k_2 = \sigma(\sigma(k_1))$ and $\sigma(k_1) \neq k_2$. Continuing this process we must eventually come to a term $a_{k_q k_1}$ where $k_1 = \sigma(k_q)$. If any off-diagonal factors remain, the process can be repeated, leading to a decomposition of the off-diagonal factors in each term into disjoint (non-touching) loops. (It is clear that if there are non-diagonal factors in a term, there must be at least two.)

All the off-diagonal elements in $\mathbf{A}$ have negative signs, so if a loop has an even number of factors their signs contribute a net positive sign to the determinant term, and a loop with an odd number of factors contributes a net negative sign. A loop with an even number of factors requires an odd number of transpositions to order the second subscripts, and *vice versa*, so a loop with an even number of factors contributes a net negative sign to $\text{sgn}(\sigma)$, and a loop with an odd number of factors contributes a net positive sign. Multiplying these two contributions, a single loop always contributes a negative sign to its determinant term. The sign contributions of disjoint loops in the same term multiply each other, so the final result is that a term with an odd number of loops is negative, and a term with an even number of loops is positive.

Recalling that an mth-order loop in the flowgraph is a set of m non-touching (disjoint) first order loops, it is now clear that (1.7), the Cramer's rule denominator in (1.6), is exactly equivalent to $1 + \sum_m (-1)^m L_m(s)$, the denominator in Mason's rule (1.3).

After transposing, the numerator in (1.6) is

$$\det\left[\mathbf{A}_1 \dots \mathbf{A}_{n-1}\ \mathbf{b}\right]^{\mathrm{T}} = \det\begin{bmatrix} 1 & -\mathfrak{T}_{12}(s) & \dots & -\mathfrak{T}_{1,n}(s) \\ -\mathfrak{T}_{21}(s) & 1 & \dots & -\mathfrak{T}_{2,n}(s) \\ \vdots & & & \vdots \\ -\mathfrak{T}_{n-1,1}(s) & -\mathfrak{T}_{n-1,2}(s) & \dots & -\mathfrak{T}_{n-1,n}(s) \\ 1 & 0 & \dots & 0 \end{bmatrix} \tag{1.9}$$

Note the contrast with the denominator determinant—there the only possible factor a_{kn} in a nonzero term was $a_{nn} = 1$, so no path or loop could include the nth state. Here (using α to denote elements of this matrix) every nonzero term must have as a factor $\alpha_{n1} = 1$; thus no other factor α_{k1} appears, so there are no real loops including state 1 in any term (the factor α_{n1} might be said to create a pseudo-loop in any $1 \to n$ path). In addition, every nonzero term must include $\alpha_{1\sigma(1)}$ for $\sigma(1) \in \{2, \dots, n\}$ and (since $\sigma(1)$ cannot occur as second subscript in any other factor in the same term) there must be a factor $\alpha_{\sigma(1)k_2}$ with $k_2 = \sigma(\sigma(1)) \neq \sigma(1)$. Since n is reachable from every state and $\alpha_{1\sigma(1)}\alpha_{\sigma(1)k_2}\dots$ includes no loops, this process must continue to result in $\alpha_{1\sigma(1)}\alpha_{\sigma(1)k_2}\dots\alpha_{k_m n}$ (after reordering the factors), i.e., there must be a path from 1 to n. Suppose this path includes an odd number of factors. Adding the pseudo-loop factor α_{n1} makes the number of factors even; since all the factors are off-diagonal, ordering the second subscripts requires an odd number of transpositions so the $1 \to n$ path contributes a negative sign to $\mathrm{sgn}(\sigma)$. Since α_{n1} is $+1$, the path contains an odd number of factors $-\mathfrak{T}_{ij}(s)$, with a negative product, and the net contribution to the term is positive. Similarly, it is easy to show that the contribution of a $1 \to n$ path with an even number of factors is always positive; thus every term in the numerator determinant contains the transmittance for some $1 \to n$ path, with a positive sign.

It may be that there is exactly one $1 \to n$ path that includes every state in

$\{1, \ldots, n\}$, with transmittance $\mathfrak{T}(s)$, in which case the numerator in (1.6) is just $\mathfrak{T}(s)$. Suppose there are m $1 \to n$ paths where, say, the kth path consists of states $\{1, k_1, k_2, \ldots, n\}$ with transmittance $\mathfrak{T}_k(s)$. There will be a term in the determinant expansion which contains, for every $j \notin \{1, k_1, k_2, \ldots, n\}$, the factor α_{jj} ; thus the value of that term will be $\mathfrak{T}_k(s)$. Since 1 must occur as first subscript in exactly one factor, and the same for n as second subscript, there cannot be more than one $1 \to n$ path in a single term; however, if the remaining states do not occur in diagonal factors α_{jj} they may form loops (disjoint from the $1 \to n$ path), in which case the same reasoning used in evaluating the signs of loops in the denominator applies. It follows that the value of the numerator determinant in (1.6) is $\sum_k \left\{\mathfrak{T}_k(s) \left[1 + \sum_m (-1)^m L_m^k(s)\right]\right\}$, Where $\mathfrak{T}_k(s)$ is the transmittance of the kth $1 \to n$ path and $L_m^k(s)$ is the sum of transmittances of all mth-order loops disjoint from the kth path. This is identical with the numerator in Mason's rule (1.3).

Thus we have shown that the solution to (1.5) yields a value for x_n, the $1 \to n$ first passage transmittance, that is the same as the result from Mason's rule.

For a clear explanation of the terms in Mason's rule, with examples, see (Phillips & Harbor 1991, Section 2.4). Mason's graphical derivation is in (Mason 1956). Engineering-oriented algebraic derivations are found in many sources, particularly (Lorens 1964, Chapter 3) and (Chen 1967). More modern (and rigorous) statistically-oriented derivations are in (Butler & Huzurbazar 2000, Section 6), (Butler 2000, Section 9), and (Butler 2001, Section 5). (Huzurbazar 2005a) has many examples of the use of Mason's rule in solving statistical flowgraphs. Linear algebra results used here can be found in any standard text, e.g., (Meyer 2000).

1.4.4 Equivalence to Pyke's solution

Pyke (1961b, Theorem 4.2) provided the following solution for first passage distributions in a Markov renewal (semi-Markov) process:

$$\mathfrak{G}(s) = \mathfrak{T}(s)\left[\mathbf{I} - \mathfrak{T}(s)\right]^{-1}\left\{\left(\left[\mathbf{I} - \mathfrak{T}(s)\right]^{-1}\right)_d\right\}^{-1} \tag{1.10}$$

where $\mathfrak{G}(s)$ is the elementwise transform of $\mathbf{G}(\tau)$, the matrix of first passage distributions, and $\mathbf{M}_d$ is the matrix with diagonal elements equal to the corresponding elements of $\mathbf{M}$ and off-diagonal elements zero.

We show that the Cramer's rule solution (1.6) for x_n, defined in Section 1.4.1 as the transform of the $1 \to n$ first passage, is equal to $\mathfrak{G}(s)_{1n} = \mathcal{G}_{1n}(s)$ as determined by (1.10). The proof is based on (Butler & Huzurbazar 2000, Section 6) and (Butler 2001, Section 5).

For brevity, let $\Delta(s) = \det\left[\mathbf{I} - \mathfrak{T}(s)\right]$ and $\Delta_{ij}(s) = (-1)^{i+j}\det\left[\mathbf{I} - \mathfrak{T}(s)\right]_{ij}$, the ijth cofactor of $\left[\mathbf{I} - \mathfrak{T}(s)\right]$, where $\left[\mathbf{M}\right]_{ij}$ is $\mathbf{M}$ with the ith row and jth column removed.

As was shown in Section 1.4.2, the denominator in (1.6) is equal to $\Delta_{nn}(s)$. By expanding the determinant of the transposed Cramer's rule numerator (1.9) about the last row, it is easy to see that the result is $\Delta_{n1}(s)$. Thus the Cramer's rule solution is

$$x_n = \frac{\Delta_{n1}(s)}{\Delta_{nn}(s)}. \tag{1.11}$$

Now using the adjugate formula for matrix inversion (Meyer 2000, Section 6.2),

$$\left[\mathbf{I} - \mathfrak{T}(s)\right]^{-1} = \frac{\left[\Delta_{ij}(s)\right]}{\Delta(s)}, \tag{1.12}$$

where $[\Delta_{ij}(s)]$ is the adjugate, the matrix of cofactors. Then

$$\left([\mathbf{I} - \mathfrak{T}(s)]^{-1}\right)_d = \frac{1}{\Delta(s)} \begin{bmatrix} \Delta_{11}(s) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \Delta_{nn}(s) \end{bmatrix} \tag{1.13}$$

so

$$\left\{\left([\mathbf{I} - \mathfrak{T}(s)]^{-1}\right)_d\right\}^{-1} = \Delta(s) \begin{bmatrix} \frac{1}{\Delta_{11}(s)} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \frac{1}{\Delta_{nn}(s)} \end{bmatrix}. \tag{1.14}$$

Now substituting (1.12) and (1.14) into (1.10) and using the fact that $\Delta(s) = \Delta_{nn}(s)$,

$$\begin{aligned} \mathfrak{G}(s) &= \mathfrak{T}(s)\,[\Delta_{ij}(s)] \begin{bmatrix} \frac{1}{\Delta_{11}(s)} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \frac{1}{\Delta_{nn}(s)} \end{bmatrix} \\ &= \mathfrak{T}(s) \begin{bmatrix} \frac{\Delta_{11}(s)}{\Delta_{11}(s)} & \frac{\Delta_{12}(s)}{\Delta_{22}(s)} & \cdots & \frac{\Delta_{1n}(s)}{\Delta_{nn}(s)} \\ \frac{\Delta_{21}(s)}{\Delta_{11}(s)} & \frac{\Delta_{22}(s)}{\Delta_{22}(s)} & \cdots & \frac{\Delta_{2n}(s)}{\Delta_{nn}(s)} \\ \vdots & & & \vdots \\ \frac{\Delta_{n1}(s)}{\Delta_{11}(s)} & \frac{\Delta_{n2}(s)}{\Delta_{22}(s)} & \cdots & \frac{\Delta_{nn}(s)}{\Delta_{nn}(s)} \end{bmatrix}. \end{aligned} \tag{1.15}$$

The $1n$th element of (1.15) is the inner product of the first row of $\mathfrak{T}(s)$ and the nth column of the second matrix factor:

$$\mathcal{G}_{1n}(s) = [0 \;\; \mathfrak{T}_{12}(s) \ldots \mathfrak{T}_{1n}(s)] \begin{bmatrix} \frac{\Delta_{1n}(s)}{\Delta_{nn}(s)} \\ \vdots \\ \frac{\Delta_{n-1,n}(s)}{\Delta_{nn}(s)} \end{bmatrix} = \frac{1}{\Delta_{nn}(s)} \sum_{k=2}^{n} \mathfrak{T}_{1k}(s)\Delta_{kn}(s). \tag{1.16}$$

Let $[\mathbf{I} - \mathfrak{T}(s)]'$ be $[\mathbf{I} - \mathfrak{T}(s)]$ (see equation 1.8) with the nth row replaced by a copy of the first row; thus $\det[\mathbf{I} - \mathfrak{T}(s)]' = 0$. Then expanding the determinant about the nth row,

$$\det[\mathbf{I} - \mathfrak{T}(s)]' = 0 = 1\Delta_{n1}(s) - \mathfrak{T}_{12}(s)\Delta_{n2}(s) - \ldots - \mathfrak{T}_{1n}(s)\Delta_{nn}(s)$$

$$\text{so} \quad \Delta_{n1}(s) = \sum_{k=2}^{n} \mathfrak{T}_{1k}(s)\Delta_{nk}. \tag{1.17}$$

It follows from the identity of the sums in 1.16 and 1.17 that

$$\mathcal{G}_{1n}(s) = \frac{\Delta_{n1}(s)}{\Delta_{nn}(s)},$$

proving the equivalence of Pyke's solution to the solution (1.11) given by Cramer's rule, and therefore to the Mason's rule solution.

1.4.5 Basing transforms on the distribution function

In some situations, the assumption that a continuous density exists is unrealistic—for example, suppose a transition represents failure of a component, and there is a nonzero probability that the component fails immediately when it is installed, or when the system is started; conditional on the component not failing immediately, it has an exponential(λ) failure distribution. Thus the holding time distribution (time to failure) has a continuous component and a point mass at 0, and the usual Lebesgue density does not exist.

In addition, the assumption of continuous densities is obviously incompatible with solving flowgraphs nonparametrically. In that case, we are given empirical distribution functions based on samples; these EDFs do have densities, but with respect to counting measure, not Lebesgue measure.

To say we can "solve" a flowgraph, i.e., find the transform of the first passage distribution between given states, asserts the existence of solutions for the system of

linear equations (1.5) presented in Section 1.4.1. The definition of the linear equations, in turn, depends on the reduction rules for series and parallel transitions given in Section 1.3. These rules are immediate consequences of properties of "transforms of distributions," which we now define.

Let $\psi(t,s)$, $s,t \in \mathbb{R}$, be a continuous kernel function, either e^{-st} (LT), e^{st} (MGF), or e^{ist} (CF). Then if $h_1(t)$ and $h_2(t)$ are arbitrary functions (not necessarily pdfs or CDFs) supported on $[0,\infty]$, by the standard mathematical definitions of transform and convolution (Rudin 1987, Section 8.13), the transform of h_1 is

$$\mathcal{T}_{h_1}(s) = \int_0^\infty \psi(t,s) h_1(t) dt$$

and the convolution of h_1 and h_2 is

$$h_1 \star h_2(x) = \int_0^\infty h_1(x-t) h_2(t) dt,$$

assuming the integrals exist. A consequence of these definitions is that

$$\mathcal{T}_{h_1 \star h_2}(s) = \mathcal{T}_{h_1}(s) \mathcal{T}_{h_2}(s).$$

Now let X and Y be non-negative independent random variables with probability measures μ, ν, respectively, which determine distribution functions F, G, and (possibly) densities f, g. Naively applying the definitions above would lead to different expressions for the transforms of the pdf and CDF:

$$\mathcal{T}_f(s) = \int_0^\infty \psi(t,s) f(t) dt \quad \text{versus} \quad \mathcal{T}_F(s) = \int_0^\infty \psi(t,s) F(t) dt.$$

However, the second integral does not have the necessary convergence properties or support the properties we need for probability distributions.

The correct form for the transform of the *distribution* is obtained by defining the transform as an expected value:

$$\mathcal{T}_X(s) = \mathrm{E}\left[\psi(X,s)\right] = \int \psi(x,s) d\mu = \int_0^\infty \psi(x,s) dF(t) = \int_0^\infty \psi(t,s) f(t) dt, \qquad (1.18)$$

where only the last integral depends on existence of a Lebesgue density. Since a distribution function is nondecreasing and bounded above by 1, it has at worst a countable number of jump discontinuities (Rudin 1976, Theorem 4.30); it follows that the next-to-last (Riemann-Stieltjes) integral always exists for some value of the parameter s (Widder 1946, Chapter I). Thus the transform of the distribution is the same whether based on the pdf or CDF.

If the transform converges in a neighborhood $-\delta < s < \delta$, $\delta > 0$ it defines a probability distribution uniquely up to sets of measure zero (Billingsley 1979, Section 30). The CF converges for $s \in (-\infty, \infty)$ because $|\exp(ist)| \leq 1$. The MGF (and therefore the LT) may or may not converge for s in an open neighborhood of 0 (e.g., it does not for the lognormal distribution). We will see in Section 2.1 that the condition always holds for transforms of EDFs.

Given this definition of the transform, the appropriate convolution operator for CDFs (based on having the required mathematical and probabilistic properties) is

$$F \star G(z) = \int_0^\infty F(z-y)dG(y) = \int_0^\infty G(z-x)dF(x).$$

See (Rosenthal 2000, Section 9.4) for a careful derivation and proof of the fact that $F \star G$ is the CDF of $Z = X + Y$. In general, the integral must be taken in the sense of Lebesgue-Stieltjes, since the Riemann-Stieltjes integral does not exist if for some x and t, F has a discontinuity at $x - t$ and G has a discontinuity at t. If densities exist, it is easy to show that, as expected,

$$\frac{d}{dx}F \star G(x) = f \star g(x) = \int_0^\infty f(z-y)g(y)dy.$$

Also assuming densities exist, using Fubini's theorem and the change of variable

$z = x + y$ it can be shown fairly easily that

$$\begin{aligned}\mathcal{T}_{F\star G}(s) &= \int_0^\infty e^{-sz}\frac{d}{dz}\left[\int_0^\infty F(z-y)g(y)dy\right]dz \\ &= \left[\int_0^\infty e^{-sx}f(x)dx\right]\left[\int_0^\infty e^{-sy}g(y)dy\right] \\ &= \mathcal{T}_F(s)\mathcal{T}_G(s). \end{aligned} \tag{1.19}$$

Using a more abstract version of Fubini's theorem (Saks 1937, Theorem 8.1) this can be shown (less easily) without the assumption of densities.

From (1.18), (1.19) and properties of the integral we see that transforms of CDFs satisfy

Linearity: $\mathcal{T}_{\alpha F+\beta G} = \alpha\mathcal{T}_F + \beta\mathcal{T}_G$.

Convolution: $\mathcal{T}_{F\star G} = \mathcal{T}_F\mathcal{T}_G$.

From (Billingsley 1979, Section 30), they also satisfy

Uniqueness: If $\mathcal{T}(s)$ converges in a neighborhood $-\delta < s < \delta$, $\delta > 0$, it defines a probability distribution uniquely up to sets of measure zero.

These are sufficient to show that these transforms satisfy the reduction rules in Section 1.3. Thus the assumption of densities in Section 1.3 entails no real loss of generality—we can operate with transforms of distributions as if they were the transforms of pdfs, though they may actually be transforms with respect to a Lebesgue-Stieltjes measure.

More detailed proofs of transform properties can be found in many texts on measure-theoretic probability; e.g., (Billingsley 1979), (Chung 2001) or (Rosenthal

2000). Proofs and discussion of properties of the Riemann-Stieltjes and Lebesgue-Stieltjes integrals can be found in (Saks 1937), (Widder 1946), or in a more expository fashion in (Burk 2007).

Though the theory above is palatable, operating with transforms of non-smooth functions can create computational difficulties; in particular, as we will see in Section 2.2.2, it complicates numerical transform inversion. An alternative is to develop smooth approximations to distributions with point masses. It is easy to see that a scalar multiple of the pdf of a Normal($\mu, \frac{1}{n}$) density, $\alpha\sqrt{\frac{n}{2\pi}}\exp\left[\frac{(t-\mu)^2}{2n}\right]$, converges to a point with probability mass α at μ as $n \to \infty$; by using such functions a distribution with continuous and discrete components (or only discrete components, such as an EDF) that is supported on $(-\infty, \infty)$ can be approximated with arbitrary precision by a continuous distribution. (For distributions supported on $[0, \infty)$, a function that is zero to the left of the origin must be used.) This is similar to the operation of kernel smoothing (Silverman 1986) applied to a set of sample points, and will be revisited in Section 2.2 when we discuss empirical mass functions.

1.5 Overview of nonparametric flowgraph models

Approximate solutions for the flowgraph models discussed in Section 1.3 can be found without assuming parametric families for holding time distributions. This is done by replacing parametric transforms, which are exact expectations of the transform kernel with respect to parametric distributions, with empirical versions, which are sample averages of the transform kernel. Empirical transforms can also be viewed as integral transforms with respect to the EDF; by substituting the Kaplan-Meier estimator for the EDF, empirical transforms can then be based on sample data subject to censoring.

Having developed empirical versions of the transforms, the remainder of the methodology is essentially unchanged, since empirical transforms can be combined using Mason's rule and the resulting transform inverted numerically (though the details of the inversion present added difficulties).

Figure 1.6 shows an overview comparing the parametric and nonparametric approaches.

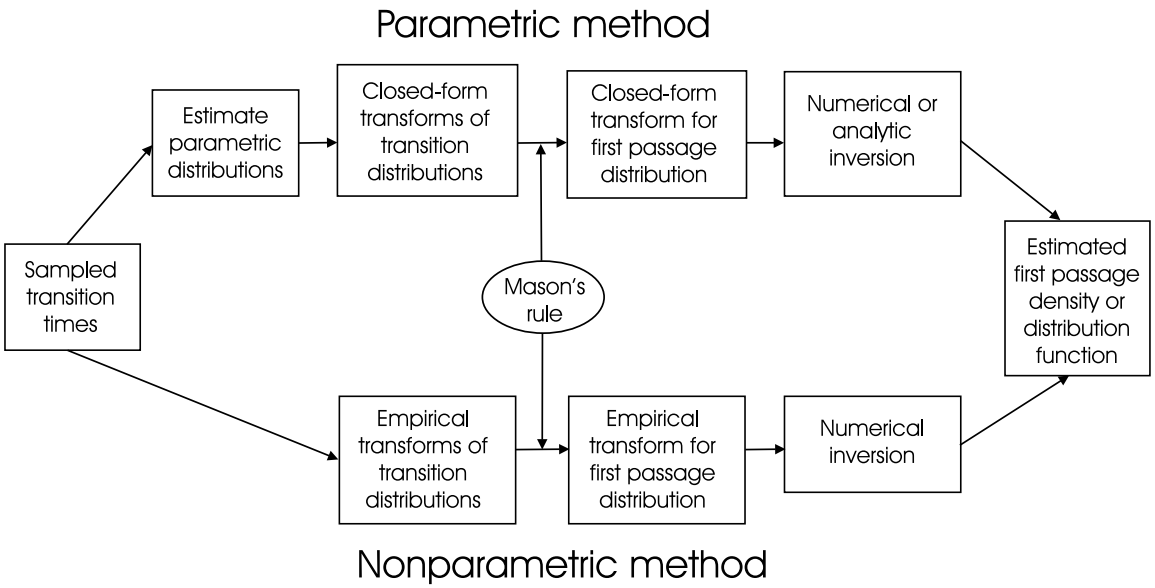


Figure 1.6: Comparison of parametric and nonparametric flowgraph modeling

A variation on this theme is semiparametric analysis. If we assume parametric distributions for some transitions and not others, the flow of the nonparametric method allows the two types of transforms to be mixed in the Mason's rule computation. This may produce a more accurate result and reduce the computation time.

There are several difficulties with the nonparametric method. One is computational: the complexity of the computation to invert the first-passage transform may

be several orders of magnitude greater than in the parametric case. Another is the accuracy of numerical estimates based on small samples. There are also a number of inferential issues such as consistency of the nonparametric estimate and the ability to develop confidence bounds for the estimate. These are all explored in Chapter 2.

Chapter 2

Methods and Analysis

This Chapter is the core of the dissertation, containing descriptions of methods we developed for nonparametric flowgraphs, analysis of results, and proofs supporting the methods. Computer code used for these methods is described in Appendix A.

2.1 Empirical transforms

Empirical transforms were briefly mentioned in Section 1.5. Their properties will be explored in detail here, and some new results proved that are important for the problem at hand.

We will call a transform defined as in (1.18), based on a member of a parametric family of distribution or density functions expressed in closed form, an *exact* or *parametric* transform. Given a random sample (the *basis sample*) $T_1, \ldots, T_n$ with T_i distributed according to F, the *empirical* (or *sample*) transform corresponding to the exact transform $\mathcal{T}$ with kernel ψ is

$$\tilde{\mathcal{T}}_F(s) = \mathrm{E}_n\left[\psi(T, s)\right] = \int_{[0,\infty)} \psi(t, s)\, dF_n(t) = \frac{1}{n}\sum_{i=1}^{n} \psi(T_i, s), \tag{2.1}$$

where F_n is the empirical distribution function (EDF) of the basis sample. We will show below that $\tilde{\mathcal{T}}_F$ has the linearity, convergence and convolution properties that one expects of an integral transform. Note that $\tilde{\mathcal{T}}_F$ is the exact transform of the EDF of its basis sample.

As an example, corresponding to the Laplace transform

$$\mathcal{L}(s) = \mathrm{E}[e^{-sT}] = \int_{[0,\infty)} e^{-st}\, dF(t)$$

is the empirical Laplace transform (ELT)

$$\tilde{\mathcal{L}}(s) = \mathrm{E}_n[e^{-sT}] = \int_{[0,\infty)} e^{-st}\, dF_n(t) = \frac{1}{n}\sum_{i=1}^{n} e^{-st_i}.$$

An absolutely precise notation would include n as a subscript or argument in expressions like $\tilde{\mathcal{T}}(s)$ to indicate the sample size on which the empirical transform is based. To avoid a confusing plethora of symbols attached to $\tilde{\mathcal{T}}$ we usually leave it out where the sample size is unimportant or evident from the context.

Empirical versions of the CF (Parzen 1962; Feuerverger & Mureika 1977), MGF (Csörgő 1982) and LT (Csörgő 1990) have been studied. An important point in the literature is that by the strong law of large numbers, the empirical transforms converge uniformly (in s) and almost surely to the transforms of the sampled distribution as $n \to \infty$. The literature also provides large sample results for the error in approximation, but only for samples drawn directly from a distribution; these are not useful for flowgraphs, where the resultant empirical transform is a complicated function of samples from several holding time distributions.

The tradeoff between parametric and empirical transforms is illustrated by an evaluation of two scenarios for estimating the MGF based on sample data (Gbur and Collins 1989): hypothesize a parametric density model $f(t;\theta)$, estimate θ as $\hat{\theta}$ using maximum likelihood, then estimate the MGF as $\hat{M}(s) = \int e^{st} f(t;\hat{\theta})\, dt$; or estimate the MGF as the empirical moment generating function (EMGF) $\tilde{M}(s) =$

$n^{-1}\sum_{i=1}^{n} e^{st_i}$. The first method (measured by mean squared error) was superior where the correct parametric family was used—not surprising given that this specification adds significant information. In the case where an *incorrect* parametric family was used, performance was significantly improved using the EMGF. Basically, methods using empirical transforms are immune from model specification errors. (A practical example of avoiding a model error by using nonparametric flowgraph modeling is presented in Section 3.3.1.)

The next section shows that empirical transforms have the linearity, convolution, and uniqueness properties described in Section 1.4.5. The following two sections address two questions that must be answered in order to fully justify the use of empirical transforms in the solution of flowgraph models:

- Suppose we compute an exact transform $\mathcal{T}(s)$ according to Mason's rule (1.3) from the exact transforms of known holding time distributions. As shown in Section 1.4, this is the transform of the first passage distribution function $G_{ij}(t)$. Let $\tilde{\mathcal{T}}(s)$ be the corresponding transform computed from empirical transforms based on holding time samples. How does $\tilde{\mathcal{T}}(s)$ relate to a hypothetical empirical transform whose basis is a sample of times from $G_{ij}(t)$?
- Is $\tilde{\mathcal{T}}(s)$ a consistent and unbiased estimator of $\mathcal{T}(s)$?

Section 2.1.4 then extends empirical transforms to allow basis samples that are censored.

2.1.1 Existence and convolution of empirical transforms

We always make the following assumptions regarding the transform kernel: $\psi(t, s)$ is continuous in both of its arguments for real $t \in [0, \infty)$ and $s \in (-\infty, \infty)$, $\psi(t, 0) = 1$ for all $0 \leq t < \infty$, and $\psi(t_1, s)\psi(t_2, s) = \psi(t_1 + t_2, s)$. All of these conditions

are satisfied by the Fourier kernel e^{ist}, the Laplace kernel e^{-st}, and the moment generating kernel e^{st}. (In fact, if we add that ψ is differentiable in t, these conditions determine that the kernel must have the form $e^{\xi t}$ for $\xi \in \mathbb{C}$—see Hardy (1952) Section IX.213. See also the theorem cited from Lukacs on page 14.)

It follows from the kernel properties that for $0 \leq t < \infty$, $\psi(t, s)$ is bounded for s in any closed interval $[-M, M]$, $0 < M < \infty$. Thus since $\tilde{\mathfrak{T}}_F(s)$ is a finite sum, it converges on every closed interval and defines a probability distribution uniquely up to sets of measure zero, namely the EDF for the given sample (Billingsley 1979, Section 30). The required linearity properties also follow immediately from the properties of finite sums.

The next result is well-known in the sense that, as mentioned in Section 1.4.5, convolution properties of transforms based on distribution functions (including EDFs) may be shown using Lebesgue integration and Fubini's theorem. However, the following proof is direct and computational, and provides more insight into the behavior of empirical transforms.

Lemma 2.1.1 (Convolution) Given non-negative RVs X and Y with CDFs F and G, and $Z = X+Y$, suppose $\{x_1, \ldots, x_n\}$ is a random sample from F and $\{y_1, \ldots, y_m\}$ is a random sample from G, with corresponding EDFs

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{[0,x]}(x_i), \qquad G_m(y) = \frac{1}{m} \sum_{j=1}^{m} I_{[0,y]}(y_i).$$

Let $\tilde{\mathfrak{T}}_F(s)$ and $\tilde{\mathfrak{T}}_G(s)$ be the corresponding empirical transforms, and $\tilde{\mathfrak{T}}_{F \star G}(s)$ the transform of the convolution $F_n \star G_m(z)$. Then $\tilde{\mathfrak{T}}_{F \star G}(s) = \tilde{\mathfrak{T}}_F(s) \tilde{\mathfrak{T}}_G(s)$.

Proof: The convolution of the EDFs is

$$\begin{aligned}
F_n \star G_m(z) &= \int_{[0,\infty)} F_n(z-y)dG_m(y) \\
&= \int_{[0,z)} F_n(z-y)dG_m(y) \qquad \text{because } F_n(z-y)=0 \text{ for } z-y<0 \\
&= \frac{1}{nm}\int_{[0,z)} \left[\sum_{i=1}^{n} I_{[0,z-y]}(x_i)\right] I_{\{y_1,\ldots,y_m\}}(y) \\
&= \frac{1}{nm}\sum_{i=1}^{n}\int_{[0,z)} I_{[0,z-y]}(x_i) I_{\{y_1,\ldots,y_m\}}(y) \\
&= \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} I_{[0,z-y_j]}(x_i) \\
&= \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} I_{[0,z]}(x_i+y_j).
\end{aligned}$$

The interchange of summation and integration is justified by the finiteness of the sums. The last line, as expected, is the EDF for $Z = X + Y$ based on the set of all possible sums $x_i + y_j$.

Now since the Lebesgue-Stieltjes measure associated with $F_n \star G_m$ is a point mass of $1/nm$ at every point $x_i + y_j$, the transform of the convolution is

$$\begin{aligned}
\tilde{\mathcal{T}}_{F\star G}(s) &= \mathcal{T}_{F_n\star G_m}(s) \\
&= \int_{[0,\infty)} \psi(z,s)\, d\left[F_n \star G_m(z)\right] \\
&= \int_{[0,\infty)} \psi(z,s)\frac{1}{nm} I_{\{x_i+y_j\}}(z) \\
&\qquad\qquad \text{where } \{x_i+y_j\} \equiv \{x_1+y_1,\ldots,x_1+y_m,\ldots,x_n+y_m\} \\
&= \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} \psi(x_i+y_j,s) \\
&= \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} \psi(x_i,s)\psi(y_j,s)
\end{aligned}$$

$$
\begin{aligned}
&= \left(\frac{1}{n}\sum_{i=1}^{n}\psi(x_i,s)\right)\left(\frac{1}{m}\sum_{j=1}^{m}\psi(y_j,s)\right)\\
&= \tilde{\mathfrak{T}}_F(s)\tilde{\mathfrak{T}}_G(s)
\end{aligned}
$$

thus proving that the transform of a convolution of EDFs is the product of the transforms of the EDFs. ∎

Lemma 2.1.5 in the next section proves a related result, that any finite product of empirical transforms $\tilde{\mathfrak{T}}_i(s)$ is the empirical transform of the convolution of the basis samples for the $\tilde{\mathfrak{T}}_i(s)$.

2.1.2 First passage sample theorem

In this section we prove an important theorem regarding the construction of sample paths for first passages in flowgraphs. Despite the length of the proof its conclusion is intuitively plausible, so we provide an overview here.

Suppose we are interested in the first passage from i to j. Let $\mathcal{R}_{ij}$ be the set of *relevant states* for the passage; i.e., $i, j \in \mathcal{R}_{ij}$, and $k \in \mathcal{R}_{ij}$ if there exists a path $i \to \cdots \to k \to \cdots \to j$ of nonzero probability. If the flowgraph has feedback loops, the set $\mathfrak{R}_{ij}$ of all such paths will contain paths in which transitions that are part of loops may occur an arbitrary number of times; thus in general $\mathfrak{R}_{ij}$ is countably infinite. For example, for the repairable system flowgraph in Figure 1.1 (page 3), $\mathfrak{R}_{1\,3}$ contains paths $1 \to 3$, $1 \to 2 \to 3$, $1 \to 2 \to 1 \to 2 \to 3$, $1 \to 2 \to 1 \to 2 \to 1 \to 2 \to 3$, etc.

Further suppose we have a sample of holding times for every adjacent-state transition in any path in $\mathfrak{R}_{ij}$. A first passage *sample path* corresponding to a path in $\mathfrak{R}_{ij}$ is an ordered collection $\{t_{ik_1}, \ldots, t_{k_nj}\}$ with each $t_{k_mk_{m+1}}$ being a sample holding

time for the corresponding $k_m \to k_{m+1}$ transition in the relevant path (again, this collection may be arbitrarily large). Then $t_{ij} = \sum_{k_m} t_{k_m k_{m+1}}$ is a sample first passage time from i to j. In principle, we could sample from $\mathfrak{R}_{ij}$ according to the distribution of path probabilities, then sample the $t_{k_m k_{m+1}}$ from the adjacent-state holding time samples, in order to construct a bootstrap sample of first passage times.

Now given that we can plug empirical transforms into Mason's rule to compute a first-passage transmittance, is this transmittance also an empirical transform? The theorem answers this question affirmatively: it is an empirical transform based on a bootstrap sample like the one just described. Aside from providing a nice logical consistency, this bears on the question of what kind of statistical inferences we can make using nonparametric flowgraph solutions.

Proving the theorem involves several steps:

- Lemma 2.1.2: Sums and products of transforms, and quotients of the form $1/[1 - \mathfrak{T}(s)]$, are transforms.

- Lemma 2.1.3: Any Mason's rule expression based on transforms of distributions is equal to the transform of a distribution. From the proof in Section 1.4.3, we know that if the base transforms are transforms of distributions for holding times between adjacent states in all paths from i to j, then the result is the transform of the first passage time distribution $G_{ij}(t)$.

- Lemma 2.1.4 and Corollary 2.1.1: A mixture of empirical transforms (corresponding to transitions in parallel) can be approximated with arbitrary precision by a single empirical transform based on a sample constructed from holding time samples for the parallel transitions. The same conclusion holds for any linear combination of empirical transforms with positive coefficients.

- Lemma 2.1.5: A product of empirical transforms (corresponding to transitions

in series) is exactly equal to a single empirical transform based on a sample constructed from holding time samples for the serial transitions.

- Lemma 2.1.6: A quotient of the form $1/[1-\tilde{\mathcal{T}}(s)]$ representing a flowgraph loop can be approximated with arbitrary precision by a single empirical transform based on a sample constructed from samples of holding times for the loop transitions.

- Theorem 2.1.1: The result of any Mason's rule computation on empirical transforms can be approximated with arbitrary precision by an empirical transform based on a bootstrap sample constructed from the holding time samples on which the individual transforms are based.

Note that Lemmas 2.1.2 and 2.1.3 are true for exact as well as empirical transforms; the remaining results are applicable only to empirical transforms.

For all these results "transform" unqualified will mean "transform of the same type as the component transforms." "Transform" unqualified will also be the transform of any function, versus "transform of a distribution," which will mean the transform of a distribution function or its density.

Recall from Section 2.1.1 (p. 38) that we assume the following properties of the transform kernel: $\psi(t,s)$ is continuous in both of its arguments for real $t \in [0,\infty)$ and $s \in (-\infty,\infty)$, $\psi(t,0) = 1$ for all $0 \leq t < \infty$, and $\psi(t_1,s)\psi(t_2,s) = \psi(t_1+t_2,s)$. As a consequence, the transforms possess the linearity and convolution properties.

It is also assumed throughout that the argument s is, first of all, restricted to an interval (γ_1,γ_2), $\gamma_1 < 0 < \gamma_2$, in which all the transforms of interest converge. It is shown below that the interval can be further restricted so that for $0 \leq p < 1$, $|p\,\mathcal{T}(s)| < 1$ for $s \in (\gamma_1,\gamma_2)$. As will be seen in Section 2.2, for practical purposes this restriction on s is without consequence, since the values of the transform that

we need for numerical inversion lie within the restricted interval.

A typical example for a term $1/[1 - p\,\mathcal{T}(s)]$, which occurs in Mason's rule computations, is shown in Figure 2.1; $p = .5$ and $\mathcal{T}(s) = 1/(1+3s)^2$, $s > -1/3$, is the LT of a gamma(2,3) distribution. On the left is a plot of $p\,\mathcal{T}(s)$, which reaches 1 at approximately -0.0976. On the right is a plot of $1/[1 - p\,\mathcal{T}(s)]$, with a pole at approximately -0.0976. In this case, we can take (γ_1, γ_2) to be, e.g., $(-.0975, \infty)$. This is a typical interval for the LT; for the MGF, the plot is reflected in the ordinate, the pole is to the right of the origin, and $\gamma_1 = -\infty$.

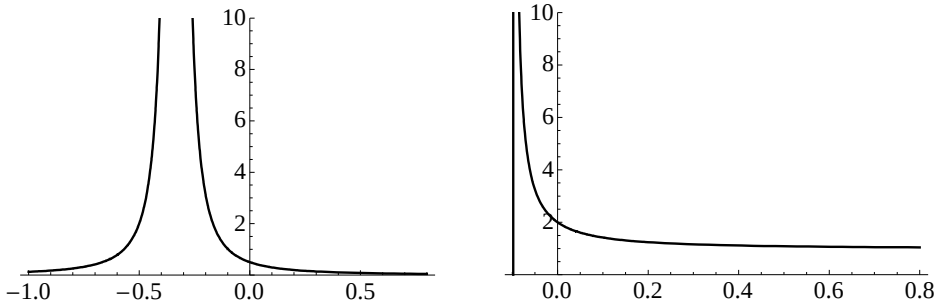


Figure 2.1: Convergence intervals

Lemma 2.1.2 (Transform properties) Suppose $\mathcal{T}_1(s), \mathcal{T}_2(s), \ldots, \mathcal{T}_n(s)$, $n < \infty$, are transforms of distributions. Then for $s \in (\gamma_1, \gamma_2)$ as defined in part (1),

1. There is an interval (γ_1, γ_2), $\gamma_1 < 0 < \gamma_2$, where for all of the i, if $0 \le p < 1$, then $|p\,\mathcal{T}_i(s)| < 1$ for $s \in (\gamma_1, \gamma_2)$.
2. For $\alpha_i \in \mathbb{R}$, $\sum_i \alpha_i \mathcal{T}_i(s)$ is a transform.
3. $\prod_i \mathcal{T}_i(s)$ is the transform of a distribution.
4. The constant 1 is the transform of a distribution.
5. For $0 < p < 1$, $1/[1 - p\,\mathcal{T}_i(s)]$ is a transform.

This Lemma holds for the CF, the LT and the MGF, as well as their empirical versions.

Proof:

1. $|\mathcal{T}_i(0)| \leq \int_{[0,\infty)} |\psi(t,0)| dF(t) = 1$ since $|\psi(t,0)| = 1$ by assumption, so $|p\,\mathcal{T}_i(0)| < 1$; the result for each i and an interval $(\gamma_{i_1}, \gamma_{i_2})$, $\gamma_{i_1} < 0 < \gamma_{i_2}$ follows by continuity. Then take $(\gamma_1, \gamma_2) = \cap_{i=1}^{n} (\gamma_{i_1}, \gamma_{i_2})$, which is an interval with length > 0 by the finiteness of n.

2. For each α_i, if $\mathcal{T}_i$ is the transform of f_i then $\alpha_i \mathcal{T}_i(s)$ is the transform of $\alpha_i f_i$; the result then follows from the linearity of the transform.

3. This follows by recursive application of the convolution property, and the convolution theorem for sums of random variables (Casella & Berger 2002, Theorem 5.2.9).

4. $\int \psi(t,s)\delta(t)dt = 1$, where $\delta(t)dt$ is Dirac measure at 0, so 1 is the transform of the distribution P_0 with a point mass of 1 at the origin. In particular, since a sample from P_0 must be $\{0, 0, \ldots, 0\}$, the empirical transform is $\tilde{\mathcal{T}}(s) = n^{-1} \sum_{i=1}^{n} \psi(0,s) = 1$.

5. By (1) $|p\,\mathcal{T}_i(s)| < 1$ for $s \in (\gamma_1, \gamma_2)$, therefore $1/[1 - p\,\mathcal{T}_i(s)]$ can be expanded in a uniformly and absolutely convergent geometric series $\sum_{i=0}^{\infty} [p\,\mathcal{T}_i(s)]^i$. Let the partial sums be $f_n(s) = \sum_{i=0}^{n} [p\,\mathcal{T}_i(s)]^i$. By (2) and (3), each f_n is a transform; since the series converges uniformly, $f_n(s) \to f(s)$, a continuous function, as $n \to \infty$. Then for the CF it follows by the well-known continuity theorem (Lukacs 1960, Theorem 3.6.1) that $f(s)$ is a CF; for the LT and MGF, the result follows by the analogous continuity theorem for Laplace-Stieltjes transforms (Feller 1971, Chapter XIII, Theorem 2a). ∎

Henceforth we always implicitly assume $s \in (\gamma_1, \gamma_2)$, as defined in part (1); e.g., if an inequality is asserted for an expression depending on s, it is asserted for all $s \in (\gamma_1, \gamma_2)$.

Lemma 2.1.3 (Mason's rule equals a transform) Any Mason's rule expression is equal to the transform of a distribution function.

Proof: From Section 1.3 we recall Mason's rule:

$$\mathfrak{T}(s) = \frac{\sum_k \left\{ \mathfrak{T}_k(s) \left[1 + \sum_m (-1)^m L_m^{\bar{k}}(s) \right] \right\}}{1 + \sum_m (-1)^m L_m(s)}. \tag{2.2}$$

$\mathfrak{T}(s)$ is the first-passage transmittance between i and j, the states of interest. $\mathfrak{T}_k(s)$ is the transmittance of the kth direct path (path with no loops) from i to j. $L_m^{\bar{k}}(s)$ in the numerator is the sum of transmittances of all mth-order loops *not* touching the kth direct path (i.e., having no states in common with it). A first-order loop is a closed path with the same start and end states that does not pass through any state more than once. An mth-order loop consists of m first-order loops with no points in common; the transmittance of an mth-order loop is the product of the transmittances of the first-order loops. $L_m(s)$ in the denominator is the sum of transmittance over all mth-order loops. Note that every $L_m^{\bar{k}}(s)$ in the numerator also occurs in the denominator, since $L_m(s)$ includes *every* mth-order loop.

Dropping the argument s for brevity and letting $\mathfrak{T}_{Lj}$ be the transmittance of the jth first-order loop, the denominator can be expanded as

$$1-(\mathfrak{T}_{L1}+\mathfrak{T}_{L2}+\mathfrak{T}_{L3}+\cdots)+(\mathfrak{T}_{L1}\mathfrak{T}_{L2}+\mathfrak{T}_{L1}\mathfrak{T}_{L3}+\mathfrak{T}_{L2}\mathfrak{T}_{L3}+\cdots)-(\mathfrak{T}_{L1}\mathfrak{T}_{L2}\mathfrak{T}_{L3}+\cdots)+\cdots$$

which factors into

$$(1-\mathfrak{T}_{L1})(1-\mathfrak{T}_{L2})(1-\mathfrak{T}_{L3})\cdots$$

Letting $\mathfrak{T}^{\bar{k}}_{Lj}$ be the transmittance of the jth first-order loop not touching the kth path, the kth term in the Mason's rule numerator summation can be expanded as

$$\begin{aligned}\mathfrak{T}_k[1-(\mathfrak{T}^{\bar{k}}_{L1}+\mathfrak{T}^{\bar{k}}_{L2}+\mathfrak{T}^{\bar{k}}_{L3}+\cdots)+(\mathfrak{T}^{\bar{k}}_{L1}\mathfrak{T}^{\bar{k}}_{L2}+\mathfrak{T}^{\bar{k}}_{L1}\mathfrak{T}^{\bar{k}}_{L3}+\mathfrak{T}^{\bar{k}}_{L2}\mathfrak{T}^{\bar{k}}_{L3}+\cdots)\\ -(\mathfrak{T}^{\bar{k}}_{L1}\mathfrak{T}^{\bar{k}}_{L2}\mathfrak{T}^{\bar{k}}_{L3}+\cdots)+\cdots]\\ =\mathfrak{T}_k(1-\mathfrak{T}^{\bar{k}}_{L1})(1-\mathfrak{T}^{\bar{k}}_{L2})(1-\mathfrak{T}^{\bar{k}}_{L3})\cdots\end{aligned}$$

Since every factor $1-\mathfrak{T}^{\bar{k}}_{Lj}$ in the numerator occurs also as a factor in the denominator, it can be canceled, and (2.2) reduces to a sum of the form

$$\mathfrak{T}(s)=\sum_k \frac{\mathfrak{T}_k(s)}{\prod_j[1-\mathfrak{T}^k_{Lj}(s)]}. \tag{2.3}$$

where $\mathfrak{T}^k_{Lj}(s)$ is the the transmittance of the jth first-order loop that *does* touch the kth path.

Since a loop does not pass through any state more than once, it consists of transitions in series and its transmittance is a product of transforms of distributions multiplied by probabilities $0 \le p_i \le 1$, thus by Lemma 2.1.2 every $\mathfrak{T}^k_{Lj}(s)$ is $p\,\mathcal{T}(s)$, the transform of a distribution multiplied by a number $p=\prod_i p_i \le 1$. By assumption, the probability of exiting the loop is nonzero so at least one of the $p_i < 1$. Then $p<1$, and by (5) of Lemma 2.1.2 the factors $1/[1-\mathfrak{T}^k_{Lj}(s)]$ in (2.3) are transforms $p^k_{Lj}\mathcal{T}^k_{Lj}(s)$, thus (2.3) becomes

$$\mathfrak{T}(s)=\sum_k \mathfrak{T}_k(s)\prod_j p^k_{Lj}\mathcal{T}^k_{Lj}(s). \tag{2.4}$$

A path consists of transitions in series, so its transmittance is also the product of transforms of distributions and probabilities $p_i \le 1$, thus by Lemma 2.1.2 each $\mathfrak{T}_k(s)$ is the transform of a distribution multiplied by a number ≤ 1. Then applying Lemma 2.1.2 from the inside out to (2.4), it is clear that every term in the product

is the transform of a distribution multiplied by a number ≤ 1, and likewise for the outer summation, so $\mathfrak{T}(s)$ is the transform of a distribution multiplied by a number $\alpha \leq 1$.

To prove the Lemma it suffices to show that $\alpha = 1$. Since (2.4) holds for $s \in (\gamma_1, \gamma_2)$ it holds for $s = 0$ and we have

$$\alpha = \sum_k p_k \prod_j p_{Lj}^k \tag{2.5}$$

where p_k is the probability of entering the kth path from i to j and p_{Lj}^k is the probability of traversing the jth loop touching the kth path. This reduces the problem to the simpler but equivalent problem of computing probabilities in the embedded Markov chain of the semi-Markov process (Çinlar 1975, Chapter 10). By the way the first passage problem is set up (see Section 1.4.1) j is an absorbing state and every other state in $\mathcal{R}_{ij}$ is transient. Since (2.5) accounts for all paths from i to j, it follows by a standard result in Markov chains (Kemeny & Snell 1960, Theorem 3.1.1) that the total probability over all the paths is $\alpha = 1$. ∎

Lemma 2.1.4 (Mixtures of empirical transforms) A finite mixture of empirical transforms based on independent samples can be approximated with arbitrary precision by an empirical transform based on a single sample from the corresponding mixture distribution. More precisely, suppose $S_1 = \{t_1^1, \ldots, t_{n_1}^1\}$, $\ldots$, $S_k = \{t_1^k, \ldots, t_{n_k}^k\}$ are random samples from distributions $F^1, \ldots, F^k$. Based on each sample construct the empirical transform with kernel $\psi(t, s)$

$$\tilde{\mathfrak{T}}_j(s) = \int_{[0,\infty)} \psi(t, s)\, dF_{n_j}^j(t) = \frac{1}{n_j} \sum_{i=1}^{n_j} \psi(t_i^j, s), \qquad j = 1, \ldots, k$$

where $F_{n_j}^j$ is the EDF based on the jth sample. Let

$$\tilde{\mathfrak{T}}(s) = \sum_{i=1}^{k} p_i\, \tilde{\mathfrak{T}}_i(s) \qquad 0 < p_i < 1,\ \sum_i p_i = 1.$$

(If $p_i = 0$ it can be dropped from the summation, and the Lemma is trivially true if any $p_i = 1$.) Then given $\epsilon > 0$ there exists a single sample $S^*_\epsilon = \{t^*_1, \ldots, t^*_{n_\epsilon}\}$, constructed from the S_i, such that

$$\left|\tilde{\mathfrak{T}}(s) - \tilde{\mathfrak{T}}^*(s)\right| = \left|\tilde{\mathfrak{T}}(s) - \frac{1}{n_\epsilon}\sum_{i=1}^{n_\epsilon} \psi(t^*_i, s)\right| < \epsilon. \tag{2.6}$$

Furthermore, S^*_ϵ is a sample (but not a random sample) from the mixture distribution $\sum_{i=1}^k p_i F^k$.

Proof: By the linearity properties of the Riemann-Stieltjes integral (Widder 1946, Theorem 5a)

$$\begin{aligned}\sum_{i=1}^k p_i\, \tilde{\mathfrak{T}}_i(s) &= \sum_{i=1}^k \int_{[0,\infty)} \psi(t,s)\, d\left[p_i F^i_{n_i}(t)\right] \\ &= \int_{[0,\infty)} \psi(t,s)\, d\left[\sum_{i=1}^k p_i F^i_{n_i}(t)\right].\end{aligned} \tag{2.7}$$

By definition of a mixture distribution, a sample from $F(t) = \sum_{i=1}^k p_i F^i(t)$ consists of elements t_i sampled from $F^1, \ldots, F^k$ with probability $p_1, \ldots, p_k$. $\tilde{F}(t) = \sum_{i=1}^k p_i F^i_{n_i}(t)$ is also a mixture distribution (based on EDFs), so a sample from it consists of elements t^*_i resampled from the samples $S^1, \ldots, S^k$ with probability $p_1, \ldots, p_k$. If we could construct a sample S^* with the property that a random element t^*_i had *exactly* the probability p_j of being drawn from S_j, then the lemma would be proved with a zero approximation error in (2.6). We will show instead that the approximation error can be made arbitrarily small by making the errors in the sampling probabilities arbitrarily small.

To construct S^*_ϵ we seek integers N_i with the property that by combining N_i copies of each S_i,

$$\frac{N_1 n_1}{\sum_{i=1}^k N_i n_i} \approx p_1, \ \ldots\ , \ \frac{N_k n_k}{\sum_{i=1}^k N_i n_i} \approx p_k. \tag{2.8}$$

By assuming equality in (2.8) a system of linear equations is obtained for for the N_j:

$$\begin{aligned} N_j n_j &= p_j \sum_{i=1}^{k} N_i n_i \qquad j = 1, \ldots, k \\ 0 &= N_1 n_1 p_1 + \ldots - N_j n_j (1 - p_j) + \ldots N_k n_k p_k \end{aligned} \tag{2.9}$$

with coefficient matrix

$$\mathrm{M} = \begin{bmatrix} -n_1(1-p_1) & n_2 p_2 & \ldots & n_k p_k \\ n_1 p_1 & -n_2(1-p_2) & \ldots & n_k p_k \\ \vdots & & & \vdots \\ n_1 p_1 & n_2 p_2 & \ldots & -n_k(1-p_k) \end{bmatrix}.$$

By a standard property of determinants,

$$\det(\mathrm{M}) = n_1 n_2 \cdots n_k \det \begin{bmatrix} -(1-p_1) & p_2 & \ldots & p_k \\ p_1 & -(1-p_2) & \ldots & p_k \\ \vdots & & & \vdots \\ p_1 & p_2 & \ldots & -(1-p_k) \end{bmatrix}.$$

By applying elementary row operations and using the fact that $\sum_i p_i = 1$, this last matrix can be transformed to one whose last row is all zeros, so $\det(\mathrm{M}) = 0$. Since the system (2.9) is homogeneous and $\mathrm{rank}(\mathrm{M}) < k$, it possesses infinitely many solutions for the N_j. In particular, given any solution $N_1 = \nu_1, \ldots, N_k = \nu_k$, $N_1 = \lambda\nu_1, \ldots, N_k = \lambda\nu_k$ is also a solution for any $\lambda \in \mathbb{R}$.

Of course, there is no guarantee that (2.9) has any solution in integers, so in general $\nu_1, \ldots, \nu_k$ is not a feasible set of replication factors for the samples S_i. To obtain integer replication factors, we use $\lfloor \lambda\nu_1 \rfloor, \ldots, \lfloor \lambda\nu_k \rfloor$, where $\lfloor \lambda\nu_i \rfloor$ is the greatest integer less than or equal to $\lambda\nu_i$ and λ is an integer large enough to make the error

small in the approximation

$$p_j \approx \tilde{p}_j = \frac{\lfloor \lambda \nu_j \rfloor n_j}{\sum_{i=1}^{k} \lfloor \lambda \nu_i \rfloor n_i}. \tag{2.10}$$

Using the fact that $\lambda \nu_j n_j - n_j \leq \lfloor \lambda \nu_j \rfloor n_j \leq \lambda \nu_j n_j$, so $\sum_i \lambda \nu_i n_i - k\, n_{\max} \leq \sum_i \lfloor \lambda \nu_i \rfloor n_i \leq \sum_i \lambda \nu_i n_i$, where $n_{\max} = \max_i(n_i)$, the approximation is bounded by

$$\begin{aligned}
\frac{\lambda \nu_j n_j - n_j}{\sum_i \lambda \nu_i n_i} &\leq \tilde{p}_j \leq \frac{\lambda \nu_j n_j}{\sum_i \lambda \nu_i n_i - k\, n_{\max}} \\
\frac{\lambda \nu_j n_j}{\sum_i \lambda \nu_i n_i} - \frac{n_j}{\sum_i \lambda \nu_i n_i} &\leq \tilde{p}_j \leq \frac{\lambda \nu_j n_j}{\sum_i \lambda \nu_i n_i} + \frac{k\, n_{\max} \lambda \nu_j n_j}{(\sum_i \lambda \nu_i n_i)\,(\sum_i \lambda \nu_i n_i - k\, n_{\max})} \\
p_j - \frac{n_j}{\sum_i \lambda \nu_i n_i} &\leq \tilde{p}_j \leq p_j + \frac{k\, n_{\max} p_j}{\sum_i \lambda \nu_i n_i - k\, n_{\max}}.
\end{aligned}$$

It is clear that the fractions on the left and right of the last line go to 0 as $\lambda \to \infty$, so the error $|p_j - \tilde{p}_j|$ can be made as small as desired by taking λ large enough.

Now let $N_i = \lfloor \lambda \nu_i \rfloor$ and construct the sample S_ϵ^* of points t_i^* as described above, leaving aside for the moment the question of a lower bound on λ that will satisfy (2.6) for a given ϵ. By construction we can partition S_ϵ^*, which is a sample from the distribution $\sum_{i=1}^{k} \tilde{p}_i F_{n_i}^i(t)$, into subsamples $S_\epsilon^{1*}, \ldots, S_\epsilon^{k*}$, where S_ϵ^{j*} has points t_i^{j*} and consists of N_i replications of S_j, so $S_\epsilon^* = \cup_{N_1} S_1 \cup \cdots \cup_{N_k} S_k$. Let $n_\epsilon = \sum_{i=1}^{k} N_i n_i$,

then the empirical transform based on S_ϵ^* is

$$\begin{aligned}
\tilde{\mathcal{T}}^*(s) &= \frac{1}{n_\epsilon}\sum_{i=1}^{n_\epsilon}\psi(t_i^*,s) \\
&= \frac{1}{n_\epsilon}\left\{\sum_{i=1}^{N_1 n_1}\psi(t_i^{1*},s) + \ \ldots\ + \sum_{i=1}^{N_k n_k}\psi(t_i^{k*},s)\right\} \qquad (2.11)\\
&= \frac{N_1 n_1}{n_\epsilon}\frac{1}{N_1 n_1}\sum_{i=1}^{N_1 n_1}\psi(t_i^{1*},s) + \ \ldots\ + \frac{N_k n_k}{n_\epsilon}\frac{1}{N_k n_k}\sum_{i=1}^{N_k n_k}\psi(t_i^{k*},s) \\
&= \tilde{p}_1\frac{1}{N_1 n_1}\sum_{i=1}^{N_1 n_1}\psi(t_i^{1*},s) + \ \ldots\ + \tilde{p}_k\frac{1}{N_k n_k}\sum_{i=1}^{N_k n_k}\psi(t_i^{k*},s) \\
&= \tilde{p}_1\frac{1}{N_1 n_1}N_1\sum_{i=1}^{n_1}\psi(t_i^{1},s) + \ \ldots\ + \tilde{p}_k\frac{1}{N_k n_k}N_k\sum_{i=1}^{n_k}\psi(t_i^{k},s) \qquad (2.12)\\
&= \tilde{p}_1\frac{1}{n_1}\sum_{i=1}^{n_1}\psi(t_i^{1},s) + \ \ldots\ + \tilde{p}_k\frac{1}{n_k}\sum_{i=1}^{n_k}\psi(t_i^{k},s) \\
&= \tilde{p}_1\tilde{\mathcal{T}}_1(s) + \ \ldots\ + \tilde{p}_k\tilde{\mathcal{T}}_k(s). \qquad (2.13)
\end{aligned}$$

In this series of equations (2.11) follows from the decomposition of S_ϵ^* into $S_\epsilon^{1*},\ldots,S_\epsilon^{k*}$; (2.12) follows from the decomposition of each S_ϵ^j into N_j copies of S_j, assuming that the ordering of sample points has been preserved. We now determine the error in (2.13) relative to $\tilde{\mathcal{T}}(s)$.

By assumption $|\tilde{\mathcal{T}}_i^*(s)| < 1$ for $s \in (\gamma_1, \gamma_2)$, for $i = 1,\ldots,k$. Now given $\epsilon > 0$,

choose λ in (2.10) such that $\max_i |p_i - \tilde{p}_i| < \epsilon/k$. Then from (2.13),

$$\begin{aligned}
\left|\tilde{\mathfrak{T}}^*(s) - \tilde{\mathfrak{T}}(s)\right| &= \left|\sum_{i=1}^{k} \tilde{p}_i \tilde{\mathfrak{T}}_i(s) - \sum_{i=1}^{k} p_i \tilde{\mathfrak{T}}_i(s)\right| \\
&= \left|\sum_{i=1}^{k} (\tilde{p}_i - p_i) \tilde{\mathfrak{T}}_i(s)\right| \\
&< \left|\sum_{i=1}^{k} (\tilde{p}_i - p_i)\right| \\
&\leq k \max_i |\tilde{p}_i - p_i| \\
&< \epsilon
\end{aligned}$$

which completes the proof. ∎

Of course, the construction of the sample S_ϵ^* by arbitrarily large replications of the original samples is merely an artifice to facilitate the proof. In practice, construction of a synthetic sample from a mixture distribution would be done using the standard algorithmic methods of computational statistics. E.g., given the mixture distribution function $\sum_{i=1}^{k} p_i F^i(t)$ with corresponding component samples $S_1, \ \ldots \ , S_k$, generate a uniform(0, 1) random variate u; if $0 < u \leq p_1$, resample S_1 into S^*, if $p_1 < u \leq p_1 + p_2$, resample S_2 into S^*, and so forth.

S_ϵ^* is a sample, but not a random sample, from the mixture distribution $\sum_{i=1}^{k} p_i F^k$, since the only values that can occur are those that are sums of the finite sets of values in $S_1, \ \ldots \ , S_k$. In addition, in practical situations the p_i are estimated from data or elicited from subject-matter experts, so in general $\sum_{i=1}^{k} p_i F^k$ is not the "true" mixture distribution. The theorem only asserts a bound for the approximation based on the given p_i.

Corollary 2.1.1 (Linear combinations of empirical transforms) Up to a constant multiple, the conclusion of Lemma 2.1.4 remains true for finite linear combi-

nations of empirical transforms with positive coefficients.. I.e., suppose

$$\tilde{\mathfrak{T}}(s) = \sum_{i=1}^{k} \alpha_i \tilde{\mathfrak{T}}_i(s) \qquad \alpha \in \mathbb{R}, \;\; \alpha > 0, \;\; \eta = \sum_{i=1}^{k} \alpha_i.$$

Then given $\epsilon > 0$ there exists a single sample $S_\epsilon^* = \{t_1^*, \ldots, t_{n_\epsilon}^*\}$, constructed from the S_i, such that

$$\left| \tilde{\mathfrak{T}}(s) - \eta \, \tilde{\mathfrak{T}}^*(s) \right| = \left| \tilde{\mathfrak{T}}(s) - \frac{\eta}{n_\epsilon} \sum_{i=1}^{n_\epsilon} \psi(t_i^*, s) \right| < \epsilon.$$

Proof:

$$\begin{aligned} \tilde{\mathfrak{T}}(s) &= \sum_{i=1}^{k} \alpha_i \tilde{\mathfrak{T}}_i(s) \\ &= \eta \sum_{i=1}^{k} \frac{\alpha_i}{\eta} \tilde{\mathfrak{T}}_i(s) \\ \frac{1}{\eta} \mathfrak{T}(s) &= \sum_{i=1}^{k} \frac{\alpha_i}{\eta} \tilde{\mathfrak{T}}_i(s) \end{aligned}$$

since $\alpha_i/\eta \leq 1$ and $\sum_i \alpha_i/\eta = 1$, it follows by Lemma 2.1.4 that there exists $\tilde{\mathfrak{T}}^*(s)$, based on S_ϵ^*, such that

$$\left| \frac{1}{\eta} \tilde{\mathfrak{T}}(s) - \tilde{\mathfrak{T}}^*(s) \right| < \epsilon$$

and the conclusion of the corollary follows. ∎

The next lemma proves a similar result for finite products of empirical transforms. Note that Lemma 2.1.5 provides an *exact* result, rather than an approximation.

Lemma 2.1.5 (Products of empirical transforms) Suppose, as in Lemma 2.1.4, that $\tilde{\mathfrak{T}}_j(s)$, $j = 1, \ldots, k$ are empirical transforms with kernel $\psi(t, s)$, based on random samples $S_1 = \{t_1^1, \ldots, t_{n_1}^1\}$, $\ldots$, $S_k = \{t_1^k, \ldots, t_{n_k}^k\}$ from distributions $F^1, \ldots, F^k$. Let

$\tilde{\mathcal{T}}(s) = \prod_{i=1}^{k} \tilde{\mathcal{T}}_i(s)$. Then there exists a single sample $S^* = \{t_1^*, \ldots, t_n^*\}$, constructed from the S_i, such that

$$\tilde{\mathcal{T}}^*(s) = \frac{1}{n} \sum_{i=1}^{n} \psi(t_i^*, s) = \tilde{\mathcal{T}}(s).$$

Proof:

$$\begin{aligned} \prod_{i=1}^{k} \tilde{\mathcal{T}}_i(s) &= \left(\frac{1}{n_1} \sum_{i=1}^{n_1} \psi(t_i^1, s) \right) \cdots \left(\frac{1}{n_k} \sum_{i=1}^{n_k} \psi(t_i^k, s) \right) \\ &= \frac{1}{n_1 \cdots n_k} \sum_{i_i=1}^{n_1} \cdots \sum_{i_k=1}^{n_k} \psi(t_{i_1}^i + \cdots + t_{i_k}^k, s) \\ &= \tilde{\mathcal{T}}^*(s) \end{aligned}$$

where $\tilde{\mathcal{T}}^*(s)$ is the empirical transform based on S^*, $n = n_1 \cdots n_k$ and S^* contains all possible sums $t_{i_1}^i + \cdots + t_{i_k}^k$ of one element from each of $S_1, \ldots, S_k$. ∎

For every loop such as the $1 \to 1$ loop in Figure 2.2 there will be a loop factor $1/[1 - \tilde{\mathfrak{T}}_L(s)]$ in the simplified Mason's rule expression (2.3) of Lemma 2.1.3. The next lemma shows the approximation of a loop factor by an empirical transform based on a single sample of holding times in the loop. $\tilde{\mathfrak{T}}_L(s) = p\,\tilde{\mathcal{T}}_L(s)$ where p is the probability of traversing the loop and $\tilde{\mathcal{T}}_L(s)$ is an empirical transform based on a sample of holding times in the loop (and in general, $\tilde{\mathcal{T}}_L(s)$ is the product of several transforms over adjacent-state transitions in the loop).

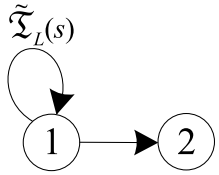


Figure 2.2: Loop transmittance

Lemma 2.1.6 (Empirical transform loop factors) Suppose that $\tilde{\mathfrak{T}}(s)$ is an empirical transform based on a random sample $S = \{t_1, \ldots, t_n\}$ from a distribution F, and $0 < p < 1$. Then given $\epsilon > 0$ there exists a positive number $\eta \in \mathbb{R}$ and an empirical transform $\tilde{\mathfrak{T}}^*(s)$ based on a single sample $S^* = \{t_1^*, \ldots, t_N^*\}$ constructed from S such that

$$\left| \eta \, \tilde{\mathfrak{T}}^*(s) - \frac{1}{1 - p \, \tilde{\mathfrak{T}}(s)} \right| < \epsilon.$$

Proof: As in part (5) of Lemma 2.1.2 the loop factor can be expanded in a convergent geometric series:

$$\frac{1}{1 - p \, \tilde{\mathfrak{T}}(s)} = 1 + p \, \tilde{\mathfrak{T}}(s) + p^2 \, \tilde{\mathfrak{T}}(s)^2 + \cdots = \sum_{i=0}^{\infty} p^i \, \tilde{\mathfrak{T}}(s)^i.$$

Let $S^0 = \{0, 0, \ldots, 0\}$ (n replications of 0), $S^1 = S$, and $S^n = S \times S \times \cdots \times S$ (n-fold Cartesian product). By part (4) of Lemma 2.1.2 the constant 1 is an empirical transform based on S^0. By hypothesis $\tilde{\mathfrak{T}}(s)$ is an empirical transform based on S^1. For $i > 1$, by Lemma 2.1.5 $\tilde{\mathfrak{T}}(s)^i$ is exactly equal to an empirical transform $\tilde{\mathfrak{T}}_i^*(s)$, which can readily be seen to be based on a sample S_i^* consisting of sums over the elements of S^i. I.e., each element of S^i is an ordered set of holding times for single traversals of the loop, so a sum over it is a sample holding time for i traversals of the loop. For $i \geq 1$, p^i is the probability of traversing the loop i times. Thus

$$\frac{1}{1 - p \, \tilde{\mathfrak{T}}(s)} = 1 + p \, \tilde{\mathfrak{T}}_1^*(s) + p^2 \, \tilde{\mathfrak{T}}_2 * (s) + \cdots = \sum_{i=0}^{\infty} p^i \, \tilde{\mathfrak{T}}_i^*(s).$$

Since the infinite sum converges absolutely there is an N such that for any $\epsilon > 0$,

$$\left| \sum_{i=0}^{\infty} p^i \, \tilde{\mathfrak{T}}_i^*(s) - \sum_{i=0}^{N} p^i \, \tilde{\mathfrak{T}}_i^*(s) \right| < \frac{\epsilon}{2}.$$

By Corollary 2.1.1 there exists a number η and an empirical transform $\tilde{\mathfrak{T}}^*(s)$ based on a single sample $S^*_\epsilon = \{t^*_1, \ldots, t^*_{n_\epsilon}\}$ constructed from the S_i, such that

$$\left| \sum_{i=0}^{N} p^i \, \tilde{\mathfrak{T}}^*_i(s) - \eta \, \tilde{\mathfrak{T}}^*(s) \right| < \frac{\epsilon}{2},$$

Then by the triangle inequality

$$\left| \sum_{i=0}^{\infty} p^i \, \tilde{\mathfrak{T}}^*_i(s) - \eta \, \tilde{\mathfrak{T}}^*(s) \right| < \epsilon, \tag{2.14}$$

which proves the theorem.

To estimate the value of η, set $s = 0$ in (2.14):

$$\left| \sum_{i=0}^{\infty} p^i - \eta \right| = \left| \frac{1}{1-p} - \eta \right| < \epsilon$$

so $\eta \approx 1/(1-p)$. ∎

Since empirical transforms are genuine transforms (of their respective EDFs), we know from Lemma 2.1.3 that if they are plugged into Mason's rule, the result is the transform of a distribution. The next Theorem shows that this is the transform of an EDF constructed from a sample of first-passage times from G_{ij}.

Theorem 2.1.1 (First passage sample theorem) Let $\mathcal{R}_{ij}$ be the set of relevant states for the passage $i \to j$. Let $\{\tilde{\mathfrak{T}}_{km}(s)\}$ be empirical transforms based on random samples $S_{km} = \{t^{km}_1, t^{km}_2, \ldots, t^{km}_{n_{km}}\}$ from transition time distributions F_{km} between adjacent states $k, m \in \mathcal{R}_{ij}$. Let $\tilde{\mathfrak{T}}(s)$ be the transmittance for first passage from state i to state j constructed according to Mason's rule from the empirical adjacent-state transmittances $p_{km}\tilde{\mathfrak{T}}_{km}(s)$, and let G_{ij} be the true $i \to j$ first-passage distribution function. Then $\tilde{\mathfrak{T}}(s) = \tilde{\mathfrak{T}}(s)$ is the transform of a distribution, and given $\epsilon > 0$ there

exists an empirical transform $\tilde{\mathcal{T}}^*(s)$ based on a sample $t_1^{ij}, t_2^{ij}, \ldots, t_N^h$ from G_{ij} such that $|\tilde{\mathfrak{T}}(s) - \tilde{\mathcal{T}}^*(s)| < \epsilon$.

Proof: The fact that $\tilde{\mathfrak{T}}(s) = \tilde{\mathcal{T}}(s)$ is the transform of a distribution follows immediately from Lemma 2.1.3. From equation (2.3) in the proof of Lemma 2.1.3 we have:

$$\tilde{\mathcal{T}}(s) = \sum_k \frac{\tilde{\mathfrak{T}}_k(s)}{\prod_j [1 - \tilde{\mathfrak{T}}_{Lj}^k(s)]} = \sum_k \tilde{\mathfrak{T}}_k(s) \prod_j \frac{1}{1 - \tilde{\mathfrak{T}}_{Lj}^k(s)}.$$

Each loop transmittance $\tilde{\mathfrak{T}}_{Lj}^k(s)$ is a product of empirical transforms multiplied by probabilities, so by Lemma 2.1.5 it is exactly equal to $p_{Lj}^k \tilde{\mathcal{T}}_{Lj}^{k*}$, where p_{Lj}^k is the product of the loop probabilities and $\tilde{\mathcal{T}}_{Lj}^{k*}$ is an empirical transform based on a sample of sums of holding times over the loop. It then follows by Lemma 2.1.6 that for $\epsilon_1 > 0$, there is an empirical transform $\tilde{\mathcal{T}}_j^{k*}(s)$ and a positive number $\eta_j^k \in \mathbb{R}$ such that

$$\left| \frac{1}{1 - \tilde{\mathfrak{T}}_{Lj}^k(s)} - \eta_j^k \tilde{\mathcal{T}}_j^{k*}(s) \right| = \left| \frac{1}{1 - p_{Lj}^k \tilde{\mathcal{T}}_{Lj}^{k*}} - \eta_j^k \tilde{\mathcal{T}}_j^{k*}(s) \right| < \frac{\epsilon_1}{2}. \tag{2.15}$$

The η's depend only on the loop probabilities, which are constant for a given flowgraph, and on ϵ_1.

Now suppose $|a_1|$, $|a_2| > \epsilon > 0$ and a_1^*, a_2^* each differ from a_1, a_2, respectively, by ϵ. Then

$$\begin{aligned} |a_1 a_2 - a_1^* a_2^*| &= |a_1 a_2 - (a_1 \pm \epsilon)(a_2 \pm \epsilon)| \\ &\leq |a_1 \epsilon| + |a_2 \epsilon| + \epsilon^2 \\ &\leq 3 \max(a_1, a_2) \epsilon \end{aligned}$$

and *a fortiori* this holds if the difference is less than ϵ. It can be readily shown by induction that under the same conditions on $a_1, a_2, \ldots, a_n$,

$$|a_1 a_2 \cdots a_n - a_1^* a_2^* \cdots a_n^*| \leq (2^n - 1) \max_i(a_i) \epsilon.$$

It follows from this and (2.15) that if the range of the product over j is J,

$$\left| \prod_{j=1}^{J} \frac{1}{1 - \tilde{\mathfrak{T}}_{Lj}^{k}(s)} - \prod_{j=1}^{J} \eta_j^k \tilde{\mathfrak{T}}_j^{k*}(s) \right| < (2^J - 1) \max_j(\eta_j^k)\epsilon_1. \tag{2.16}$$

From Lemma 2.1.5 there is a transform $\tilde{\mathfrak{T}}^{k*}(s)$ such that exact equality holds in

$$\eta^k \tilde{\mathfrak{T}}^{k*}(s) = \prod_{j=1}^{J} \eta_j^k \tilde{\mathfrak{T}}_j^{k*}(s)$$

where $\eta^k = \prod_j \eta_j^k$.

Each path transmittance $\tilde{\mathfrak{T}}(s)$ is a product of empirical transforms multiplied by probabilities, so by Lemma 2.1.5 it is exactly equal to $p_k \tilde{\mathfrak{T}}_k^*$, where p_k is the product of the path probabilities and $\tilde{\mathfrak{T}}_k^*$ is an empirical transform based on a sample of sums of holding times over the path. Then from (2.16), since $|p_k \tilde{\mathfrak{T}}_k^*| < 1$, if the range of the sum over k is K,

$$\left| \sum_k \tilde{\mathfrak{T}}_k(s) \prod_j \frac{1}{1 - \tilde{\mathfrak{T}}_{Lj}^{k}(s)} - \sum_{k=1}^{K} p_k \tilde{\mathfrak{T}}_k^*(s) \eta^k \tilde{\mathfrak{T}}^{k*}(s) \right| < K(2^J - 1) \max_{j,k}(\eta_j^k)\epsilon_1. \tag{2.17}$$

Now set $\epsilon_1 = \epsilon / 2K(2^J - 1) \max_{j,k}(\eta_j^k)$ so the righthand side of the inequality (2.17) becomes $\epsilon/2$. By Lemmas 2.1.5 and 2.1.4 there is an empirical transform $\tilde{\mathfrak{T}}^*(s)$ (the lack of a constant multiplier follows from Lemma 2.1.3) such that

$$\left| \sum_{k=1}^{K} p_k \tilde{\mathfrak{T}}_k^*(s) \eta^k \tilde{\mathfrak{T}}^{k*}(s) - \tilde{\mathfrak{T}}^*(s) \right| < \frac{\epsilon}{2}$$

so by the triangle inequality

$$\left| \sum_k \tilde{\mathfrak{T}}_k(s) \prod_j \frac{1}{1 - \tilde{\mathfrak{T}}_{Lj}^{k}(s)} - \tilde{\mathfrak{T}}^*(s) \right| < \epsilon$$

and the precision claim of the theorem is proved.

To show that the sample on which $\tilde{\mathcal{T}}^*(s)$ is based is a sample from G_{ij}, we unwind the result as follows. From Lemma 2.1.6, the approximation to each term $1/[1 - \tilde{\mathfrak{T}}^k_{Lj}(s)]$ is based on a sample of holding times for $0, 1, 2, \ldots$ traversals of the jth loop touching the kth path. From Lemma 2.1.5, the convolution represented by the product is based on the sums of all possible combinations of holding times for traversal of all the loops as the kth path is traversed. By Lemma 2.1.4, the transform of the sum is based on a sample from the mixture of holding times across all the K paths. This is $\tilde{\mathcal{T}}^*(s)$, which is therefore based on a sample of holding times for first passage through all possible paths from the start state i to the end (absorbing) state j, i.e., a sample from G_{ij}. ∎

Note that for the reasons stated after Lemma 2.1.4 the sample is not a *random* sample from G_{ij}; we can only assert that it is a *possible* sample, since it is constructed based on feasible paths and actually observed holding times for adjacent-state transitions on the paths.

2.1.3 Convergence to exact transforms

A large part of the justification for the use of empirical transforms in flowgraph modeling follows from the Lemma and Theorem below, which show that Mason's rule applied to empirical transforms provides a consistent estimator of the Mason's rule expression based on the corresponding exact transforms.

Lemma 2.1.7 (Consistency and unbiasedness of empirical transforms) Let $\mathcal{T}(s)$, with transform kernel $\psi(t, s)$, be the LT, MGF, or CF of a distribution function $F(t)$, and suppose the transform exists (converges) for $s \in [\gamma_1, \gamma_2]$, $\gamma_1 < 0 < \gamma_2$. Let $S = \{T_1, T_2, \ldots, t_n\}$ be a random sample of size n drawn from F, and $\tilde{\mathcal{T}}(s)$ the empirical transform based on S. Then $\tilde{\mathcal{T}}(s)$ is an unbiased and uniformly strongly consistent estimator of $\mathcal{T}(s)$.

Proof: Where the expectation is taken with respect to F,

$$\begin{aligned} \mathrm{E}[\tilde{\mathfrak{T}}(s)] &= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\psi(T_i,s)\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[\psi(T_i,s)] \\ &= \mathfrak{T}(s) \end{aligned}$$

by definition of the transform and because the t_i are iid F-distributed, so $\tilde{\mathfrak{T}}(s)$ is an unbiased estimator of $\mathfrak{T}(s)$.

Uniform strong consistency is equivalently stated as $n^{-1}\sum_{i=1}^{n}\psi(t_i,s) \to \mathfrak{T}(s)$ almost surely as $n \to \infty$, with the convergence being uniform over $s \in [\gamma_1, \gamma_2]$. For the LT, this is proved in Proposition 1 of (Csörgő 1990); for the MGF, in Proposition 1 of (Csörgő 1982); and for the CF, in Theorem 2.1 of (Feuerverger & Mureika 1977). ∎

Theorem 2.1.2 (Consistency of rational functions of empirical transforms) Let $\mathfrak{T}_1(s), \mathfrak{T}_2(s), \ldots, \mathfrak{T}_k(s)$ be LTs, MGFs, or CFs (all the same type) of distribution functions $F_1, F_2, \ldots, F_k$, with $(\Omega_j, \mathcal{F}_j, P_j)$ the probability space underlying infinite sequences of iid F_j-distributed RVs T_j. Suppose the T_j are mutually independent and $\{t_{j\,1}, t_{j\,2} \ldots, t_{j\,n_j}\}$ is a random sample of size n_j drawn from the jth distribution. Let $\tilde{\mathfrak{T}}_1(s), \tilde{\mathfrak{T}}_2(s), \ldots, \tilde{\mathfrak{T}}_k(s)$ be the corresponding empirical transforms based on the samples and suppose all the transforms exist for $s \in [\gamma_1, \gamma_2]$, $\gamma_1 < 0 < \gamma_2$.

If $Q(\mathfrak{T}_1(s), \ldots, \mathfrak{T}_k(s))$ is a rational function of the $\mathfrak{T}_j(s)$, then with respect to the product measure $P_1 P_2 \cdots P_k$, $Q(\tilde{\mathfrak{T}}_1(s), \ldots, \tilde{\mathfrak{T}}_k(s)) \to Q(\mathfrak{T}_1(s), \ldots, \mathfrak{T}_k(s))$ almost surely for $s \in [\gamma_1, \gamma_2]$ as $\min_j(n_j) \to \infty$. In particular, a Mason's rule solution based on empirical transforms converges a.s. to the Mason's rule solution based on the corresponding exact transforms as the minimum sample size goes to infinity.

Proof: By the mutual independence of the T_j, $P = P_1 P_2 \cdots P_k$ is a probability measure on the space $\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_k$ of collections of size k of infinite sequences, the jth sequence being iid F_j-distributed. For each j let $N_j \subset \Omega_j$ be the set of sequences on which $\tilde{\mathcal{T}}_j(s)$ does *not* converge to $\mathcal{T}_j(s)$, and $\mathbf{N}_j = \Omega_1 \times \cdots \times N_j \times \cdots \times \Omega_k$; by Lemma 2.1.7, $P_j(N_j) = 0$, so by construction $P(\mathbf{N}_j) = 0$. $\mathbf{N} = \cup_{i=1}^{k} \mathbf{N}_j$ is then the set in Ω on which at least one $\tilde{\mathcal{T}}_j(s)$ fails to converge to $\mathcal{T}_j(s)$, and $\mathbf{N}$ also has P-measure 0 since k is finite. It follows that $\mathbf{N}^c = \Omega \backslash \mathbf{N}$, the set on which $\tilde{\mathcal{T}}_j(s) \to \mathcal{T}_j(s)$ for all $j = 1, \ldots, k$, has P-measure 1.

Now if we consider only collections of sequences in $\mathbf{N}^c$, $\tilde{\mathcal{T}}_j(s) \to \mathcal{T}_j(s)$ for all $j = 1, \ldots, k$ in the ordinary pointwise sense. It follows by standard theorems of analysis that for $i,\ j \in \{1, \ldots, k\}$, $\tilde{\mathcal{T}}_i(s) + \tilde{\mathcal{T}}_j(s) \to \mathcal{T}_i(s) + \mathcal{T}_j(s)$ pointwise, $\tilde{\mathcal{T}}_i(s)\tilde{\mathcal{T}}_j(s) \to \mathcal{T}_i(s)\mathcal{T}_j(s)$ pointwise, and $\tilde{\mathcal{T}}_i(s)/\tilde{\mathcal{T}}_j(s) \to \mathcal{T}_i(s)/\mathcal{T}_j(s)$ pointwise (since the denominator is never zero). By applying these results recursively, it follows that any rational function $Q(\tilde{\mathcal{T}}_1(s), \tilde{\mathcal{T}}_2(s), \ldots, \tilde{\mathcal{T}}_k(s))$ converges pointwise to $Q(\mathcal{T}_1(s), \mathcal{T}_2(s), \ldots, \mathcal{T}_k(s))$ on $\mathbf{N}^c$, provided the denominator is nonzero, for every $s \in [\gamma_1, \gamma_2]$. Since $\mathbf{N}^c$ has P-measure 1, the general convergence result on Ω is P-almost sure. ∎

By the Markov property, the mutual independence condition of the theorem is always satisfied for holding time distributions and their transforms in semi-Markov processes.

The next logical step would be to show, using the terminology of Theorem 2.1.2, that $\tilde{Q} = Q(\tilde{\mathcal{T}}_1(s), \ldots, \tilde{\mathcal{T}}_k(s))$ is an unbiased estimator of $Q = Q(\mathcal{T}_1(s), \ldots, \mathcal{T}_k(s))$ for finite samples (it follows from consistency that it is asymptotically unbiased). We conjecture that this is true, but have not been able to prove it. Since the values of the transforms for fixed values of s are independent, it immediately follows from properties of sums and products of random variables that the numerator of $\tilde{Q}$ is an unbiased estimate of the numerator of Q, and similarly for the denomina-

tor. However, if $Q = \phi/\gamma$, it does not follow from $\mathrm{E}(\phi) = f$ and $\mathrm{E}(\gamma) = g$ that $\mathrm{E}(\phi/\gamma) = f/g$. It is true that $\mathrm{E}(\phi/\gamma) = \mathrm{E}(\phi)\mathrm{E}(1/\gamma)$, so proving unbiasedness requires a knowledge of the distribution of $1/\gamma$. Even in special cases the distribution of quotients of RVs is non-trivial, and although in principle a solution can be found (using Mellin transforms) for the general case (Springer 1979), in practical terms this may be intractable. Thus we leave this problem on the agenda for future research.

2.1.4 Empirical transforms based on censored data

In many reliability and survival applications only part of a component lifetime is observed. Dealing with such censored data is well understood for parametric flowgraph modeling, affecting the likelihood function for an assumed family of distributions (Huzurbazar 2005a; Klein and Moeschberger 2003).

For nonparametric modeling we adjust the transforms to account for censoring. We consider only random right censoring; in terms of multistate models this means that the time of entry to state i is observed, but at random, the time of transition from i to j is not observed, but rather the time of "loss to observation." More precisely, there is a competing risk between the holding time T_{ij} distributed according to $F_{ij}(t)$ and an independent censoring time T_{ij}^{C} distributed according to $C_{ij}(t)$; what is observed is $Z_{ij} = \min(T_{ij}, T_{ij}^{\mathrm{C}})$ and the value of the indicator $I_{\{T_{ij} \le T_{ij}^{\mathrm{C}}\}}$. If the indicator is 0 we know only that $T_{ij} > Z_{ij}$. Random right censoring is the type most commonly seen in practice. For example, in reliability testing a unit under test is accidentally broken by the experimenter before it fails; or in a clinical trial a subject moves away before experiencing the event of interest.

The method described below relies on an estimate of the EDF (or the empirical survival function, 1 minus the EDF). The standard for right-censored data is the product limit (PL) estimator of the survival function $\hat{S}(t)$ due to Kaplan and Meier

(1958), but variations on the PL estimator exist for other types of censoring (Klein and Moeschberger 2003, Chapters 4-5). Suffice it to say that if an estimate of the EDF is available, we can estimate empirical transforms; however, the specific results given here apply only to random right censoring and the PL estimator.

In place of the the standard product limit calculation used by Kaplan and Meier, it is more convenient for the present purposes to use an algorithm developed and shown to be equivalent to K-M by Efron (1967). This is called "redistribute to the right" and works as follows: for a censored sample of size n, initially assign a probability mass of $1/n$ to each of the ordered sample points $z_{(1)}, z_{(2)}, \ldots, z_{(n)}$, then scan the points from left to right. If $z_{(i)}$ is an uncensored observation leave it unaltered; if it is censored, remove it from the sample and distribute its mass evenly over the remaining points to its right (if the last point in the sample is censored, it it kept). The result is an estimate $\{\alpha_1, \ldots, \alpha_m\}$ of the empirical mass function (EMF), where $m \leq n$ is the number of failure (uncensored) points, $0 \leq \alpha_i \leq 1$, and $\sum_i \alpha_i = 1$.

Efron then summed over the EMF to estimate the EDF or survival function; we use it directly to estimate the empirical transform as

$$\tilde{\mathfrak{T}}^*(s) = \int_{[0,\infty)} \psi(t,s)\, d\hat{F}_n(t) = \sum_{i=1}^{m} \alpha_i \psi(t_i^*, s) \tag{2.18}$$

where the t_i^* are failure points from the original sample.

Kaplan and Meier (1958, Section 5) showed that for right-censored data, $\hat{F}_n(t) = 1 - \hat{S}_n(t)$ is the nonparametric maximum likelihood estimator (NPMLE) of the distribution function. Stute and Wang (1993, Theorem 1.1) showed that under right censoring and mild regularity conditions on the censoring mechanism, functionals $\int_{[0,\infty]} \varphi(u) d\hat{F}_n(u)$ of $\hat{F}_n$, where φ is Borel-measurable, are strongly consistent estimators of the corresponding functionals of F. In particular, by taking $\varphi(u) = I_{[0,t]}$,

$\hat{F}_n(t)$ itself is a strongly consistent estimator of $F(t)$. Gill (1980, Theorem 4.1.1) showed that in fact $\hat{F}_n(t)$ is uniformly strongly consistent. By taking $\varphi(u) = \psi(u, s)$,

$$\lim_{n\to\infty} \tilde{\mathcal{T}}_n^*(s) = \lim_{n\to\infty} \int_{[0,\infty]} \psi(u,s) d\hat{F}_n(u) = \int_{[0,\infty]} \psi(u,s) dF(u) = \mathcal{T}(s) \text{ almost surely,}$$

so the empirical transform based on the Kaplan-Meier estimator is a strongly consistent estimator of the exact transform.

Though the Kaplan-Meier estimator of the distribution function is consistent (and therefore asymptotically unbiased), it is biased downward for finite samples (Meier 1975, Section 2.2), i.e., $\mathrm{E}[\hat{F}_n(t)] \leq F(t)$. This is intuitively plausible, since the failure points $\{t_1^*, \ldots, t_m^*\}$ in a sample subject to censoring are not sampled from $F(t)$, but from the conditional distribution $F(t)|F(t) \leq C(t)$, where $C(t)$ is the censoring distribution. Both (Meier 1975) and (Gill 1980b) provide bounds on the bias of $\hat{F}_n(t)$; (Stute 1994) proves a bound for the bias of functionals $\int_{[0,\infty]} \varphi(u) d\hat{F}_n(u)$, which is sufficient to conclude that empirical transforms based on samples with censoring also have negative bias, i.e., $\mathrm{E}[\tilde{\mathcal{T}}^*(s)] \leq \mathcal{T}(s)$. The quantitative bounds are not very useful for actual data analysis, since they require knowledge of the censoring distribution that is seldom available.

By the invariance of the NPMLE (Owen 2001), it follows that $\tilde{\mathcal{T}}^*(s)$ as computed above in (2.18) is the NPMLE of $\mathcal{T}(s)$. From the strong consistency of empirical transforms based on the Kaplan-Meier estimator, we have the following corollary to Theorem 2.1.2 (p. 61):

Corollary 2.1.2 (Consistency of censored data empirical transforms) If $\tilde{\mathcal{T}}_1^*(s), \tilde{\mathcal{T}}_2^*(s), \ldots, \tilde{\mathcal{T}}_k^*(s)$ are computed from censored samples as in (2.18), the conclusion of Theorem 2.1.2 remains valid. ∎

The empirical transform constructed as in (2.18) from censored data cannot be construed as being directly based on the sample failure points $\{t_1^*, \ldots, t_m^*\}$, since the

probability mass attached to each point is not identically $1/m$. In order to be able to treat empirical transforms without regard for censoring, we would like to say that any $\tilde{\mathcal{T}}^*(s)$ constructed from a censored sample can be approximated with arbitrary precision by an empirical transform $\tilde{\mathcal{T}}(s) = N^{-1}\sum_{i=1}^{N} t_i$ based on an uncensored sample $\{t_1, \ldots, t_N\}$. This is shown by the following

Theorem 2.1.3 (First passage sample theorem for censored data) Suppose $\tilde{\mathcal{T}}^*(s) = \sum_{i=1}^{m} \alpha_i \psi(t_i^*, s)$ is an empirical transform based on a censored sample with failure points $\{t_1^*, \ldots, t_m^*\}$. Then for any $\epsilon > 0$ there exists a sample $\{t_1, \ldots, t_N\}$ such that

$$\left|\tilde{\mathcal{T}}^*(s) - \tilde{\mathcal{T}}(s)\right| = \left|\tilde{\mathcal{T}}^*(s) - \frac{1}{N}\sum_{i=1}^{N} t_i\right| \leq \epsilon.$$

Proof: We merely sketch the proof, since the details are essentially the same as in Lemma 2.1.4 (p. 48). Pick $k_1, \ldots, k_m$ such that $\sum_{i=1}^{m} k_i = N$ and $k_i/N \approx \alpha_i$; construct a sample S from k_i copies of t_i^*, so $S = \{t_1^*, \ldots, t_1^*, \ldots, t_m^*, \ldots, t_m^*\}$ where each t_i occurs k_i times. Then

$$\tilde{\mathcal{T}}(s) = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{k_i} t_{ij}^* \approx \sum_{i=1}^{m} \alpha_i t_i^* = \tilde{\mathcal{T}}^*(s). \tag{2.19}$$

See Lemma 2.1.4 for the details of bounding the approximation for α_i so that the absolute error in (2.19) is smaller than ϵ. ∎

Section 3.1.2 presents an example of empirical flowgraph analysis using censored data.

2.2 Numerical inversion of empirical transforms

Having constructed from adjacent-state holding time samples an estimate $\tilde{\mathcal{T}}(s)$ for the transform of the first passage distribution $G_{ij}(\tau)$, we now need to recover information about $G_{ij}(\tau)$ itself. Our strategy, following the parametric flowgraph methodology, is to numerically invert $\tilde{\mathcal{T}}(s)$ to obtain an approximation of the first passage density $g_{ij}(\tau)$, from which percentiles and approximations to other functions of interest such as the CDF and hazard rate can be computed.

Consistent with advice frequently given by experts on numerical methods, we validate results by using two different inversion methods: numerical Laplace transform inversion, and the saddlepoint approximation. Aside from validation, it will be seen that each method has advantages and disadvantages in terms of computational overhead and accuracy for various types of distribution.

After some general remarks we describe and provide examples for the methods we have focused on: a Fourier series approximation to the standard inversion integral for the Laplace transform, and an empirical version of the well-known saddlepoint approximation based on the moment generating function. These methods were chosen based on experience and heuristics, and there are many roads not taken in this dissertation. In the literature on transform inversion and nonparametric density estimation one will find methods and variations almost beyond counting, much of the justification for which is empirical results within the constraint of acceptable computational overhead.

Our test cases for inversion methods use simulated data, which allows comparison of density estimates to true densities. The primary global comparison will be the integrated absolute error in the density, or L_1 distance: if $\hat{f}$ is an estimate of f, the IAE is $\int_0^\infty |\hat{f}(t) - f(t)|\, dt$. A more common measure is the L_2 distance or integrated square error $\int_0^\infty [\hat{f}(t) - f(t)]^2 dt$, because of its more desirable mathematical properties

(Silverman 1986); but see (Devroye & Györfi 1985) for a contrary view. Here we only need a numerical measure of the error, and IAE is somewhat more convenient for that purpose.

For comparison of CDF approximations, we use the Kolmogorov-Smirnov (K-S) distance (or statistic) $\sup_t |\hat{F}(t) - F(t)|$. The K-S statistic can be used for a formal test of goodness-of-fit to the true distribution; we have not presented results of this test, but in all our examples the K-S statistic for the best inversion method is smaller than the critical value for a test at the 1% significance level (usually much smaller), indicating a good fit. In some cases a more localized comparison is helpful, and we use the relative error of specific percentiles of the distribution.

In addition, we place an informal value on smooth estimates, since one of the purposes of nonparametric estimation of the pdf or CDF is to allow visual assessment of fit to a known family of distributions.

Further details on the algorithms used, with computer code listings, will be found in Appendix A.

2.2.1 The general problem

Based on the first passage sample theorem (Theorem 2.1.1, p. 57) it is reasonable to begin a discussion of the general problem of density estimation as if we had the empirical transform of a single sample from a given distribution, so we consider an empirical transform $\tilde{\mathfrak{T}}(s) = n^{-1} \sum_{i=1}^{n} \psi(t_i, s)$.

The time-to-event distributions of interest in reliability and survival analysis are almost without exception of the absolutely continuous type, so we assume henceforth that we are dealing with empirical transforms based on samples from distributions having densities with respect to Lebesgue measure. In fact, even the space of dis-

tributions with rational Laplace transforms, a subset of the absolutely continuous distributions, is dense (in the sense of convergence in law) in the space of all probability distributions (Asmussen 2003, Theorem 4.2), so there is little loss of generality in this assumption. However, the assumption of a continuous density creates a mismatch with the raw material, the transform of a discrete EMF with a probability mass of $1/n$ at each sample point t_i. Thus any reasonable inversion method must either smooth the density in some way, or reproduce the sample points exactly so that something like kernel smoothing (Silverman 1986) may be applied. In practice, computational considerations mainly dictate the first alternative, though we use a variant on the second in one place (p. 83 ff.).

An alternative to the measure-theoretic treatment of point masses is the engineer's Dirac delta "function" defined by $\delta(x) = 0$ for $x \neq 0$ and $\int_{-\infty}^{\infty} \delta(x)dx = 1$. Thus a point mass of $1/n$ at t_i is formally equal to $n^{-1}\,\delta(x - t_i)$. The delta function can be handled rigorously as a measure (Bauer 2001, Section 1.3, Example 5), as a "generalized function" (Zemanian 1987, Section 1.3), or as a sequence of functions whose limit is the Dirac delta (Lang 1997, Section XI.1). The latter is the most useful here, since in numerical work the sequence must be truncated at some finite limit, which (assuming the approximating functions are smooth) results in a smooth approximation to the delta.

One possible approximation is developed as follows. Reproduce a point mass at a periodically at intervals of 2π; then, since $\delta(x)$ is integrable, we can represent it by its Fourier series; by truncating the series at a finite number k of terms and restricting the argument to $[a - \pi, a + \pi]$, we get an approximation to the delta:

$$\delta(x-a) \approx d_k(x-a) = \begin{cases} \frac{1}{2\pi} + \frac{1}{\pi}\sum_{i=1}^{k} \cos(x-a) & a-\pi \leq x \leq a+\pi \\ 0 & \text{otherwise.} \end{cases} \tag{2.20}$$

This is plotted in Figure 2.3 for $a = 5$, $k = 30$.

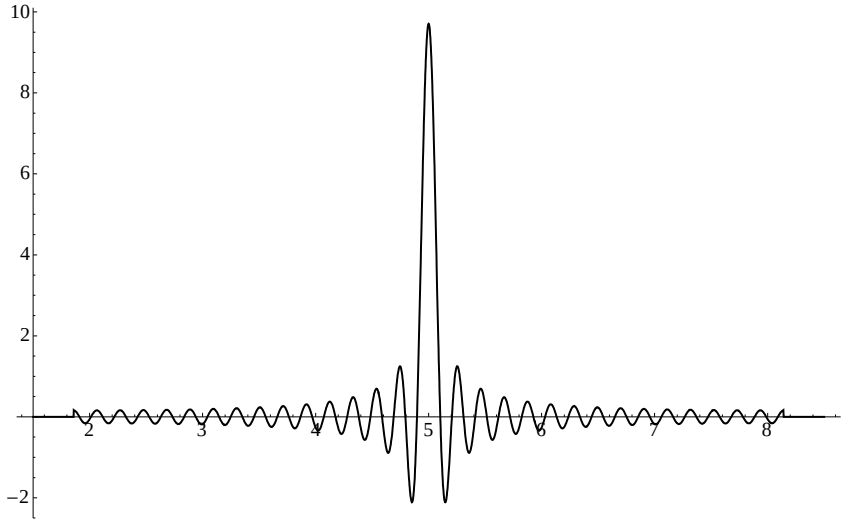


Figure 2.3: Fourier series approximation of a Dirac delta (point mass)

Though $d_k(x) \to \delta(x)$ as $k \to \infty$ and $\int_{-\infty}^{\infty} d_k(x)dx = 1$, it has the disadvantages of not being nonnegative, and for $a < \pi$, having support extending beyond the origin on the left. By using a Gaussian density function with variance k^{-1} instead of d_k the first disadvantage is avoided, and by using a gamma density with variance k^{-1} both are avoided. However, we will continue with d_k as defined above because it provides some insight into the behavior of the Fourier series approximation for inversion of the Laplace transform, to be discussed below in Section 2.2.2.

A single term $\psi(t_i, s)$ of the empirical transform $\tilde{\mathcal{T}}$ is the transform of a Dirac delta at t_i, since $\int_0^{\infty} \psi(t, s)\delta(x-t_i)dt = \psi(t_i, s)$, so trivially the transform is invertible exactly, and the inversion recovers the sample points. In this simple case, it would be appropriate to apply methods such as kernel smoothing to estimate the density $f(t)$, since the statistical properties of these methods are better understood than the properties of numerical transform inversion.

In practice, exact inversion of empirical transforms is possible only for the simplest

flowgraphs. For more complex flowgraphs such as the one illustrated in Figure 1.1 (p. 3), this is not feasible. Though in principle (by Theorem 2.1.1) the first passage transform can be approximated to arbitrary precision using a single sample, in the actual case we have only the Mason's rule solution (1.2) (p. 16), which for empirical transforms may contains tens of thousands of terms; even using a powerful computer algebra system it is not feasible to reduce this expression to something that looks like an empirical transform with a single basis sample. However, as described in the next two sections there are inversion methods that will compute a result that approximates what we would obtain from the single-sample transform.

For the sake of a simple illustration we revert to $\tilde{\mathcal{T}}(s)$ based on a sample $\{t_1, \ldots, t_n\}$. Using (2.20) the approximate inversion of $\tilde{\mathcal{T}}(s)$ is $n^{-1} \sum_{i=1} d_k(x - t_i)$, for suitably chosen k. Figure 2.4 shows an example based on a sample of size 1,000 from a gamma(2,5) density. A sample histogram is plotted along with the true density (dotted line) and the approximation with $k = 5$ (solid line).

In keeping with the assumption (which we know to be true in this case) that the density to be estimated is from the class of those typically encountered, the plethora of modes is a problem. In signal processing terms (Mallat 1999, Section 3.1.2), these reflect noise introduced by the sampling process. A simple solution under the assumption of fewer modes is to filter out high-frequency noise using a low-pass filter, which is accomplished by reducing the number k of terms in the truncated Fourier series. This is shown for varying k in Figure 2.5, where a scaled $d_k(0)$ is shown with each approximation. (In the actual computation, the height of $d_k(0)$ gets larger with increasing k, since it always integrates to 1.) The approximation is better with smaller k, a fact often noted in the literature of curve estimation using Fourier series (Tarter & Lock 1993; Efromovich 1999)—though of course "better" depends on the *a priori* assumption of a smooth curve without too many bumps.

As mentioned above, we may also approximate the Dirac delta at a with a

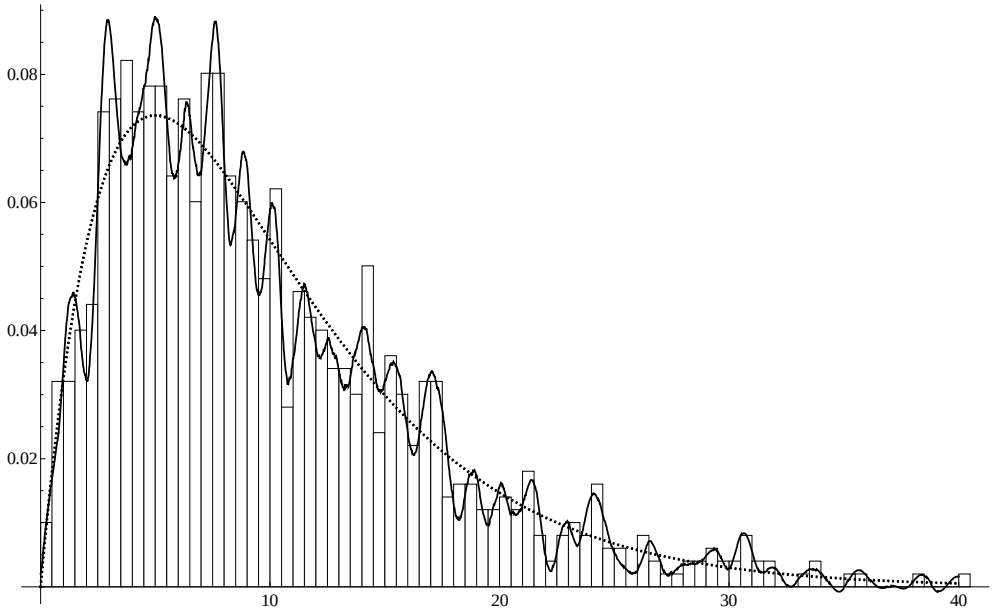


Figure 2.4: Approximation of a gamma(2,5) density using d_5

Gaussian density function $\frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-a)^2}{2\sigma^2}\right]$, where σ controls the spread of the delta. The density estimate is then

$$\frac{1}{n}\sum_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-t_i)^2}{2\sigma^2}\right], \tag{2.21}$$

which is exactly equivalent to kernel smoothing with a Gaussian kernel and bandwidth σ. Figure 2.6 shows this approximation for the same gamma sample as above, with various bandwidths.

By a well-known result for normal mixtures of normal distributions, if the t_i are iid $\mathrm{N}(\mu, \eta^2)$, then in the limit as $n \to \infty$ the mixture density 2.21 is a normal (Gaussian) density, $\mathrm{N}(\mu, \eta^2 + \sigma^2)$. If the t_i have a non-normal distribution function F, the limit distribution still has mean μ, but rather than being normal, it is the convolution of F with the Gaussian kernel. For large σ the Gaussian component

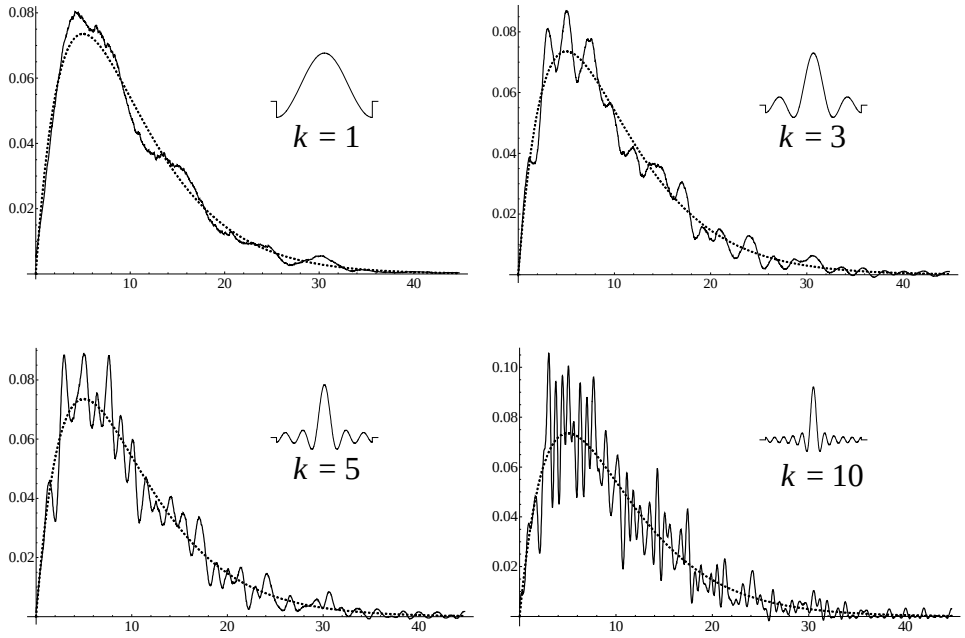


Figure 2.5: Approximation by sums of d_k (inset) with various k

dominates and the result is pulled toward a central-limit approximation, as illustrated in Figure 2.6 for $\sigma = 5$. A similar phenomenon (for a different but related reason) occurs in the saddlepoint approximation, to be discussed in Section 2.2.3. In general, Fourier-based estimates tend to be undersmoothed and oscillatory; Gaussian-based estimates are oversmoothed toward the central-limit approximation as the kernel variance increases.

Why densities?

Given an iid sample, the empirical distribution function has well-defined asymptotic properties such as convergence to the true CDF, confidence intervals for point estimates, etc. (Serfling 1980, Section 2.1). Proofs are relatively straightforward, and

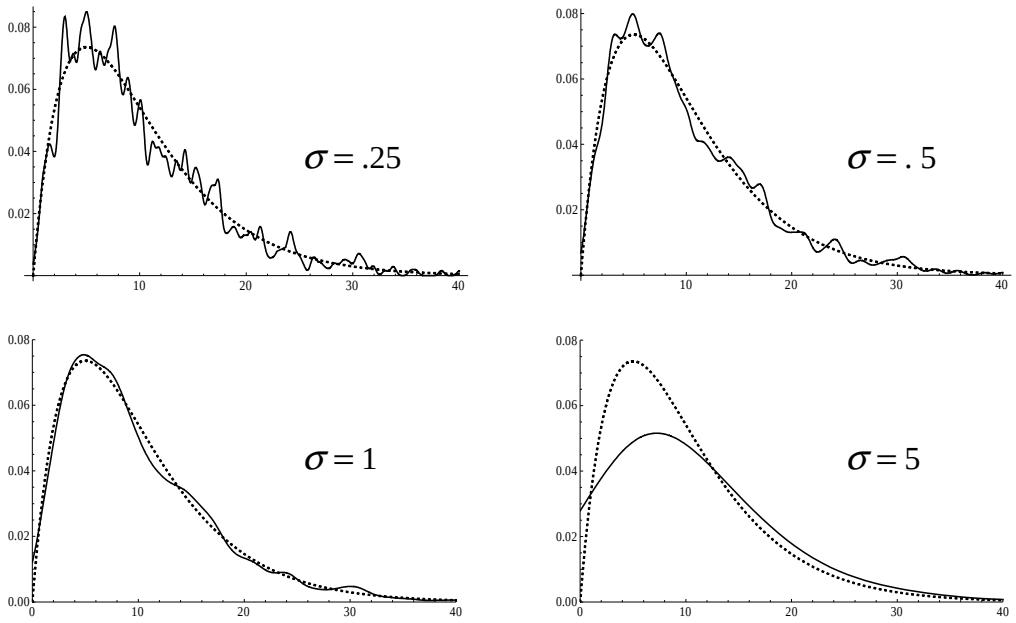


Figure 2.6: Approximation by sums of Gaussian kernels with various values of σ (the true density is the dotted plot).

rely on weak conditions. On the other hand, the vast literature on density estimation is filled with clever and difficult proofs of results that are generally weaker, method-specific (e.g., a proof of consistency only for kernel smoothing), and in some cases dependent on conditions that are impossible to verify from a sample, such as existence of higher-order derivatives of the true density; see, e.g., (Devroye & Györfi 1985; Silverman 1986; Tarter & Lock 1993; Wand & Jones 1995; Efromovich 1999).

However, for several reasons the theoretical advantages of distribution functions are hard to realize in the present context. The basis for our approximations is a Mason's rule expression which in a certain sense is based on a sample of first passage times, but not an iid sample. In addition, as will be discussed in the next several sections, no exact analysis exists for the error introduced by the inversion methods.

The tools at hand readily provide density approximations, and there is no evidence that direct inversion to a CDF approximation is more accurate; numeric integration to derive a CDF approximation from a density approximation is numerically stable, whereas the reverse process (numeric approximation of derivatives) is not. Finally, in practical multistate problems interest often centers on identification of distribution families. The popularity of density estimates such as histograms, stem-and-leaf plots, etc., suggests that there are reasons based in human perception and cognition why such devices facilitate generating and visual checking of hypotheses regarding the distribution type, more so than plots of the CDF.

The business of density estimation, by any method, can sound rather distressingly heuristic. As two experts say, "It does not seem possible to provide effective methods with simple general error bounds ... we propose using two different methods ... Assuming that the two methods agree to within desired precision, we can be confident of the computation" (Abate & Whitt 1995, p. 36). Confidence is a relative thing, but sometimes we can only make progress through heuristics, and hope that subsequent replication and analysis leads to a more rigorous foundation for our conclusions.

When we do need an estimate of the distribution function (or the survival function, or the hazard function), it can be readily computed from a density estimate; examples are shown in the next two sections.

2.2.2 Fourier series approximation

Here we discuss a method of inverting the Laplace transform which is "classical" in the sense of using no special statistical properties of functions, except insofar as they manifest themselves as straightforward mathematical properties. Laplace transform inversion is generally considered a difficult problem—see (Davies & Martin 1979;

Abate & Whitt 1992; Duffy 1993) and references therein. In a formal sense, it is intractable, since exact Laplace transform inversion $\mathcal{L}^{-1}$ is an unbounded operator (Bellman *et al.* 1966, Section 2.9); i.e., if $f^*(s) = \mathcal{L}[f(t)]$, for any $\epsilon > 0$ and $M \gg 0$ there exists a function $\psi(t)$ such that for s in an arbitrary neighborhood (γ_1, γ_2) where the transform converges, $\sup_s |\psi^*(s)| < \epsilon$ but

$$\sup_s \left| \mathcal{L}^{-1}[f^*(s) + \psi^*(s)] - \mathcal{L}^{-1}[f^*(s)] \right| > M.$$

In the context of numerical work this means that arbitrarily small errors in the transform can lead to arbitrarily large errors in the inverse. (For a different reason, the same problem of intractability will be seen in Section 2.2.3 to arise for the saddlepoint inversion.)

In principle, the situation might be even worse in the context of empirical transforms because our *desideratum* is not to produce a result which is the inverse of the transform, but rather a smooth approximation of (we hope) the unknown density from which the transform is sampled. However, in practice, for the functions of interest, namely bounded absolutely continuous probability densities with a limited number of modes (local maxima of the density), the problem becomes much more tractable than in the general case.

The method under discussion here, due to Abate and Whitt (1992, 1995), uses a Fourier series to approximate the inversion integral for the Laplace transform. Following Abate and Whitt, we refer to it as the EULER algorithm. It is optimized for inversion of pdfs of nonnegative random variables with absolutely continuous distributions, which are functions with special characteristics: supported on the positive half-line, nonnegative, bounded, continuous, and in the cases we are interested in, decaying rapidly at infinity.

The algorithm is remarkably accurate in the inversion of exact Laplace transforms of continuous densities; so good, in fact, that we do not show any examples—on a

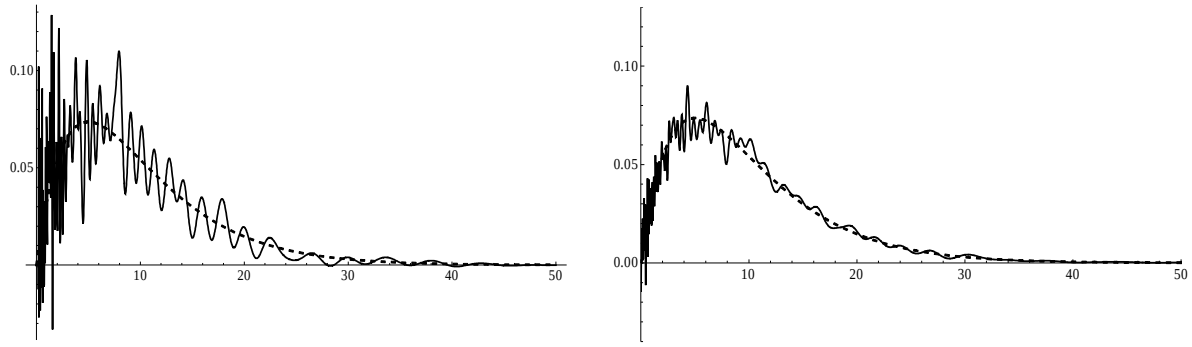


Figure 2.7: EULER inversion (solid line) of an empirical Laplace transform (left: small sample; right: large sample). The dashed line is the true density.

plot, the EULER inversion result is visually indistinguishable from the exact density in every case we have tested. For both exact and empirical transforms, our results with EULER are superior to those with other well-known algorithms we have tried, such as the Weeks method, which approximates the transformed function using a Laguerre series (Weideman 1999).

EULER will be described in more detail below, but first we display examples of its application to empirical transforms. Figure 2.7 shows on the left the inversion of an empirical transform whose basis sample is the same gamma(2,5) sample of size 1,000 shown in Figure 2.4. On the right is the inversion of the transform based on a sample of size 10,000 from the same distribution. Even with the larger sample size, the inversion is quite noisy; some methods for producing a smoother result will be discussed below. The underlying accuracy of the approximation can be seen from the distribution function approximations in Figure 2.8, computed by numeric integration of the density approximations shown in Figure 2.7; Kolmogorov-Smirnov statistics for the small and large sample CDF approximations are 0.02554 and 0.01573, respectively.

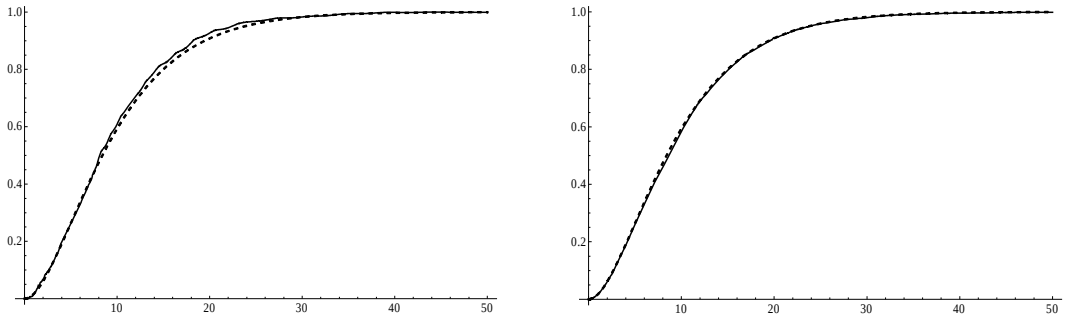


Figure 2.8: Distribution function computed from the EULER approximation of a gamma pdf (left: small sample; right: large sample). The dashed line is the true CDF.

Sample sizes on the order of 10^4, or even 10^3, are completely out of the question in most practical applications of flowgraphs. But recalling the results of Section 2.1.2 regarding the construction of sample paths in flowgraphs, the effective sample size for the first passage distribution in a flowgraph with loops and series of transitions is on the order of the product of the sizes of individual transition samples. Thus an effective sample size of 10^4 or larger for the distribution of interest is not unusual, and this is reflected in the results for flowgraphs such as the repairable system of Figure 1.1 (p. 3)—see Figure 3.3 (p. 125).

The exact transform for the gamma(2,5) density is plotted in figure 2.9 along with the empirical transform based on the sample of size 1,000, for real s in a neighborhood of the origin. Given the apparent accuracy of the empirical approximation, at first sight one might wonder "where the noise comes from" in Figure 2.7. The explanation is in Figure 2.10, which plots the empirical transform along the contour in the complex plane which is the path of integration for the Laplace transform inversion integral; on paths parallel to the imaginary axis, the empirical transform is a sum of trigonometric functions and is highly oscillatory. As will be seen in the next section,

the saddlepoint approximation uses only real arguments in the empirical transform, and has quite a different behavior.

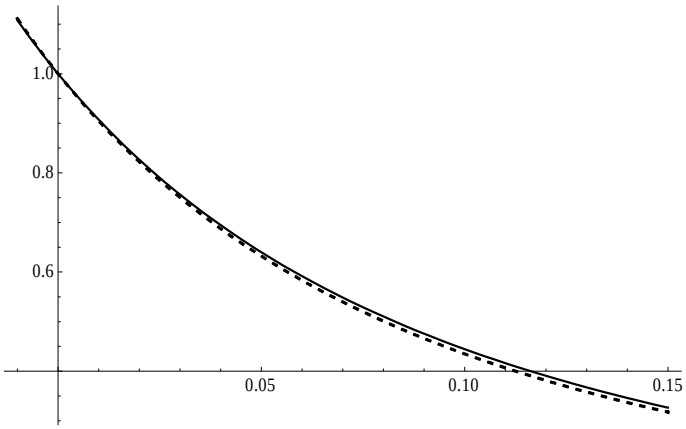


Figure 2.9: Exact transform of the gamma(2,5) density (solid line) versus empirical transform from a sample of size 1,000 (dashed line) for real s near the origin.

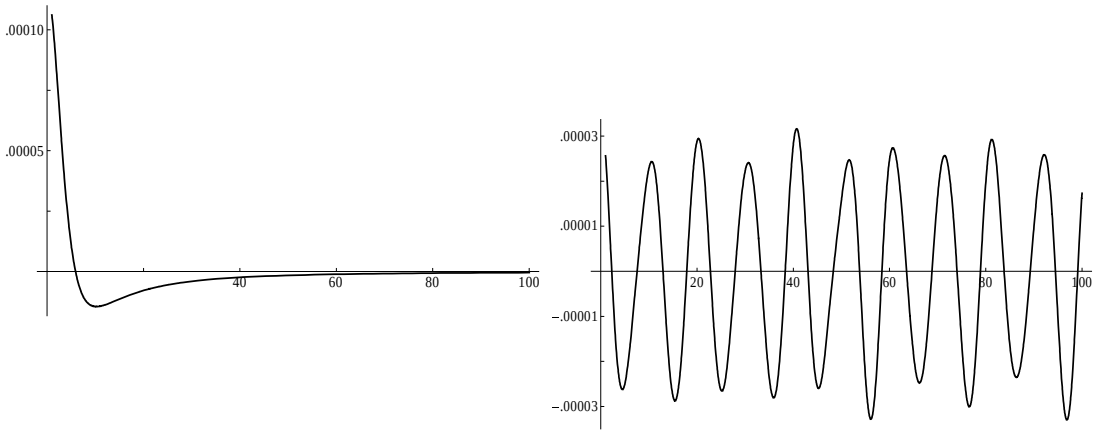


Figure 2.10: Exact transform of the gamma(2,5) density (left) versus empirical transform from a sample of size 1,000 (right) for s on the complex contour of integration used by EULER.

Derivation of the EULER algorithm

To provide some background on the Fourier series approximation and to understand how it might be improved to deal with empirical transforms, we sketch out the derivation of the EULER algorithm. The details can be found in Sections 3-6 of (Abate & Whitt 1992); (Doetsch 1974) presents the theory behind inversion using the Bromwich integral, which is what EULER approximates, and (Dubner & Abate 1968) is helpful in understanding the reduction of the approximation to a cosine series.

Where $f^*(s)$ is the Laplace transform of the pdf $f(t)$ of a nonnegative random variable, we start with the standard inversion integral (Doetsch 1974, Chapter 24); a is a point on the real axis such that $f^*(s)$ has no singularities in the complex plane on or to the right of the line $s = a$:

$$\begin{aligned} f(t) &= \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} e^{st} f^*(s)ds \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{(a+iu)t} f^*(a+iu)du \qquad \text{changing the variable } s \to a+iu \\ &= \frac{2e^{at}}{\pi}\int_0^{\infty} \mathrm{Re}\{f^*(a+iu)\cos ut\}du. \end{aligned} \tag{2.22}$$

The last line follows by using $e^x = \cos x + i \sin x$, the fact that $f(t)$ is real and vanishes for $t < 0$, and the fact that cosine is an even function.

Now the integral (2.22) is approximated using the trapezoidal rule with a subinterval size of h:

$$\begin{aligned} f(t) &\approx f_h(t) \\ &= \frac{he^{at}}{\pi}\mathrm{Re}\{f^*(a)\} + \frac{2he^{at}}{\pi}\sum_{k=1}^{\infty}\mathrm{Re}\left\{f^*(a+ikh)\cos(kht)\right\}, \end{aligned}$$

which is a Fourier cosine series. The trapezoidal rule works as well as other methods of quadrature here because of the oscillatory integrand. For given t, using $h = \pi/2t$ makes the cosine 0 for odd k and $(-1)^{k/2}$ for even k; then reindexing,

$$f_h(t) = \frac{e^{at}}{2t}\mathrm{Re}\{f^*(a)\} + \frac{e^{at}}{t}\sum_{k=1}^{\infty}(-1)^k \mathrm{Re}\left\{f^*\left(a + \frac{ik\pi}{t}\right)\right\}. \tag{2.23}$$

Subject to the condition that there are no singularities to its right (and note that the empirical Laplace transform has no singularities to the right of 0), a may be an arbitrary positive number. For $A > 0$ chosen based on a tradeoff between accuracy and ease of computation (Abate & Whitt 1992, p. 30), set $a = A/2t$ in (2.23), then

$$f_h(t) = \frac{e^{A/2}}{2t}\mathrm{Re}\left\{f^*\left(\frac{A}{2t}\right)\right\} + \frac{e^{A/2}}{t}\sum_{k=1}^{\infty}(-1)^k \mathrm{Re}\left\{f^*\left(\frac{A}{2t} + \frac{ik\pi}{t}\right)\right\}. \tag{2.24}$$

Let $s_N(t)$ be (2.24) with the sum truncated after the Nth term. This partial sum approximates $f(t)$, but as a final step the algorithm applies Euler summation to accelerate convergence:

$$f(t) \approx \sum_{m=0}^{M}\binom{M}{m}2^{-m}s_{N+m}(t). \tag{2.25}$$

$\binom{M}{m}$ are binomial coefficients, and N and M are chosen heuristically to bound the error of approximation. Our tests indicate that the Euler summation, for the same error, reduces the number of terms computed by as much as a factor of 10.

Abate and Whitt provide a partial error analysis for their algorithm, but only for functions without point masses, whereas our functions are entirely composed of point masses. Furthermore, we are not trying to recover the transformed function, but rather the function from which the point masses were sampled. The issue of error analysis is discussed further in Section 2.3.

EULER is attractive for two reasons: experience in using it for parametric flowgraphs, which result in exact transforms, indicates that it is remarkably accurate even in cases where the saddlepoint approximation performs poorly (cf. Section 2.2.3, p. 95 ff.); and for nonparametric flowgraphs, its computational overhead is significantly less than that of the empirical saddlepoint approximation.

Smoothing the Fourier series approximation

We have evaluated various alternatives for smoothing the results of inversion using EULER, of which three appear worthy of further analysis:

- Presmoothing the sample points for each transition using a standard kernel smoothing method, then transforming the results and using them as input to Mason's rule, with the final result as input to EULER.
- Exponentially smoothing the density approximation output from EULER.
- Using a modified EULER algorithm with fewer terms in the summation.

Presmoothing applies a standard kernel smoothing method to the sample points to produce a preliminary density estimate, which is then (in the flowgraph context) transformed and input to Mason's rule, or for our test cases below, used as direct input to EULER. This results in a smoother output since the input is now the transform of a continuous (though not necessarily smooth) density. Following Silverman (1986, Section 3.5), since kernel smoothing is equivalent to convolution of the data with the kernel function, this can be effected by simply multiplying the empirical transform by the Laplace transform of the kernel. (Silverman developed the method to speed up the computation of kernel density estimates by discretizing and using the fast Fourier transform.)

To make this method work effectively and efficiently for our samples ideally requires a kernel that smooths well, has a simple Laplace transform, and puts no mass to the left of the origin. The Gaussian kernel satisfies the first two conditions, but not the third. Failure to satisfy the third condition causes problems for EULER, so we have compromised on an asymmetric rectangular kernel, i.e., for bandwidth h, the kernel is a uniform$(0, h)$ density with Laplace transform $(1 - e^{-hs})/hs$. This is applied to the sample after left shifting by $h/2$ and reflection of negative points to avoid bias; see Appendix A.3.2 for details. For smoothing a single sample there are much better kernels. For example, (Chen 2000) describes an adaptive gamma kernel which works very well for the types of density we deal with in reliability and survival analysis, but is difficult to use in a flowgraph context since its Laplace transform must be recomputed for each sample point. The rectangular kernel is not a very good smoother *per se*, but does significantly improve the performance of EULER.

The top row of Figure 2.11 shows the results for the gamma(2,5) sample of 1,000 points. On the left is the histogram and the direct kernel-smoothed estimate; on the right is the output of EULER after convolving the kernel with the empirical transform of the sample points. The bottom row shows the same plots based on a presmoothed sample of 10,000 points.

Of course, given a single sample there is no point in going through the extra steps of transforming and inverting the presmoothed result (unless we use Silverman's fast-Fourier method). Kernel smoothing alone can produce a better result with less overhead using a kernel superior to the rectangular, which was adopted here only for ease of computation. In a real flowgraph problem, though, it might make sense to presmooth and transform each transition sample, combine the results using Mason's rule, and then invert.

The second alternative for improving the Fourier series inversion result is post-smoothing of the approximation, i.e., applying a smoothing algorithm to the set of

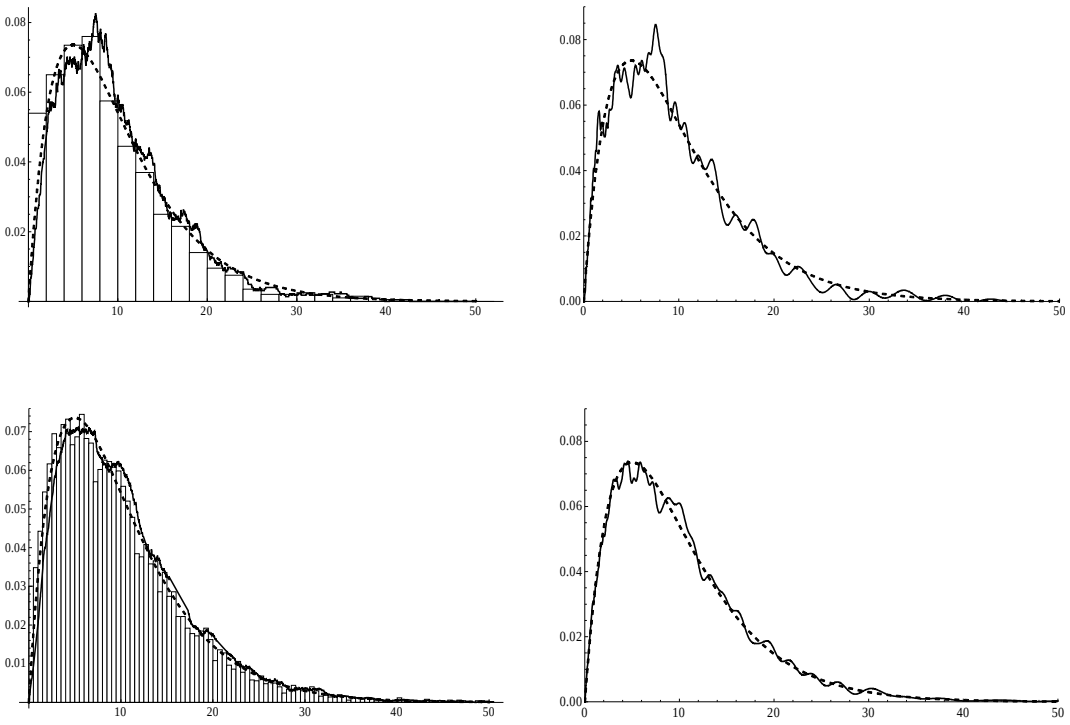


Figure 2.11: EULER inversion of the transform of a presmoothed gamma(2,5) sample (top left: histogram and kernel smoothing; top right: EULER output, n = 1,000; bottom: same plots for n = 10,000). The dashed line is the true density.

points computed as a density approximation. We have achieved our best result with an exponentially weighted moving average, which assumes that the true density at a point t is a weighted average of point approximations computed by the inversion in a neighborhood of t, with the weights falling off exponentially with distance from t. Specifically, for uniformly spaced points t_i, $i = 1, \ldots, N$, let $y_i = \hat{f}(t_i)$ be the corresponding density values computed by EULER; then the smoothed points are given by $\tilde{y}_i = \frac{\alpha}{2} \sum_{k=-n}^{n} (1-\alpha)^{|k|} y_{i+k}$, where $0 < \alpha < 1$, n is large enough that the weights sum approximately to 1, and the sums are truncated at 0 and N.

Figure 2.12 shows an example set of weights for $\alpha = .2$; by the well-known convergence of the geometric density to the exponential, as the number of points

goes to infinity the weighting converges to a Laplace (double exponential) density $\frac{\alpha}{2}e^{-\alpha|t|}$, shown as a solid line in the figure.

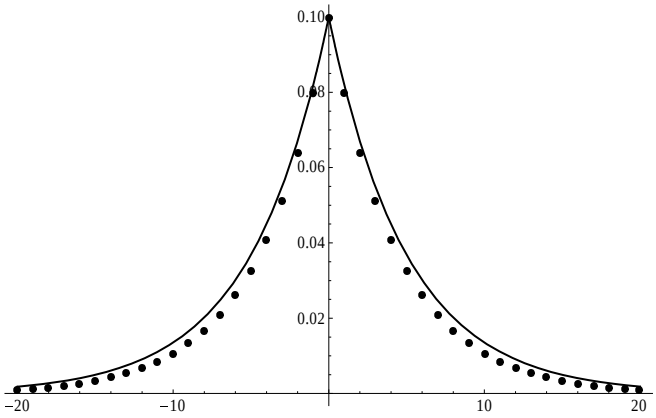


Figure 2.12: Exponential smoothing weights for $\alpha = .2$, with Laplace density (solid line).

Exponential smoothing has been extensively treated in the time series literature; see (Chatfield 2004) for references. In that context it is asymmetric and used for forecasting, whereas centered exponential smoothing is symmetric, a combination of forecasting and "backcasting;" however, the theory is essentially the same when each direction is considered separately. It is shown in (Gijbels *et al.* 1999) that exponential smoothing is equivalent to nonparametric kernel regression, which is more intuitive in this context—we base the density at each point on a weighted regression on nearby points.

Figure 2.13 shows the result of applying symmetric exponential smoothing to the Fourier inversion plot points from gamma(2,5) samples, with $\alpha = .04$. The choice of α is based on visual smoothness and the constraint that the smoothed density integrate approximately to one; the α that meets these criteria will be smaller as the mesh of points over which the density is approximated becomes finer. Typically we use several thousand plot points and α in the range .02–.05.

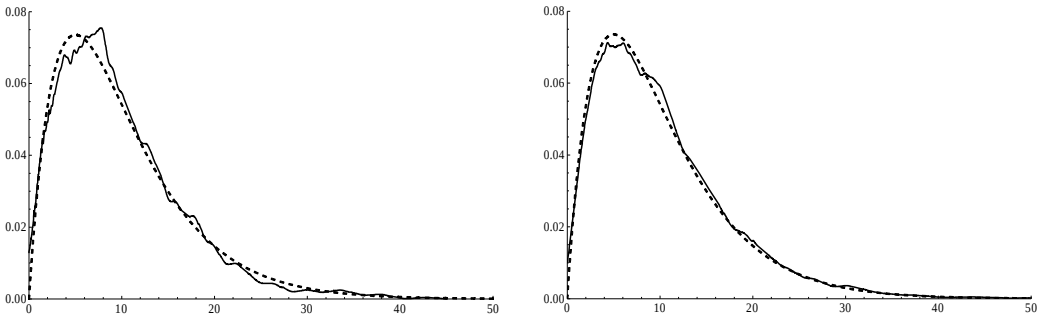


Figure 2.13: Exponentially smoothed EULER inversion of the transform of a gamma(2,5) sample. Left: n = 1,000; right: n = 10,000. The true pdf is the dashed line.

The third alternative we have examined is smoothing by reducing the number of terms in the partial Fourier sums used by EULER. In Equation (2.25) we take $N = 0$ (versus 15 in EULER) and $M = 12$ (versus 11 in EULER); this reduces the total number of terms computed for each approximated point from 27 to 14. Again, it is a heuristic procedure to determine the optimum number of terms, based on the criteria of smoothness, nonnegativity of the density (which is an issue because of the approximation by cosines), and integration to approximately 1. (In this case, because of the amount of computation, we have made the choices once and for all, whereas the exponential smoothing parameter α can be easily adjusted for each case.) Results using the modified algorithm for samples from the gamma(2,5) distribution are shown in Figure 2.14.

Table 2.1 compares the integrated absolute errors for approximating the gamma(2,5) density by these three methods.

We now consider another example, a mixture distribution; specifically $f(t) = .5f_1(t)+.5f_2(t)$ where f_1 is a gamma(10,0.5) density and f_2 is a gamma(50,3) density. As we will see in the next section, this density and others of its type are approximated

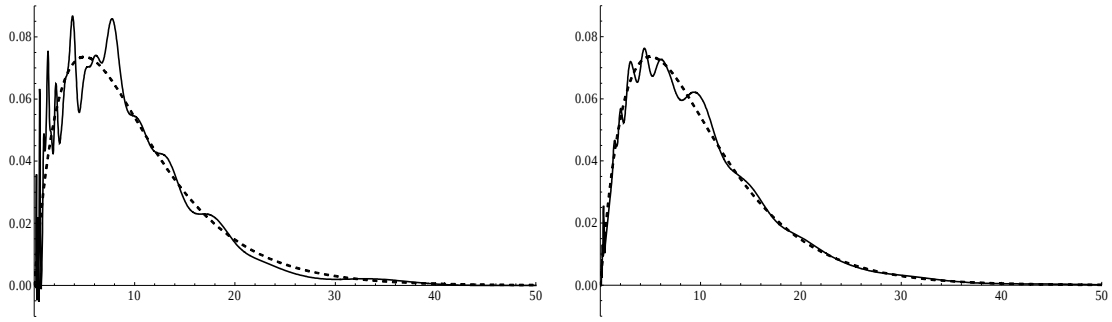


Figure 2.14: Modified EULER inversion of empirical transforms of gamma(2,5) samples (solid lines) (left: small sample; right: large sample). The dashed line is the true density.

Method	IAE	
	$n = 1{,}000$	$n = 10{,}000$
EULER	0.29167	0.09370
EULER (kernel presmoothing)	0.10899	0.05841
EULER (exponential smoothing)	0.07871	0.05076
EULER (modified algorithm)	0.12153	0.05202

Table 2.1: Integrated absolute errors (IAE) for EULER variations, gamma(2,5) samples.

much better by EULER than by the saddlepoint approximation. Figure 2.15 shows a sample from the mixture (500 observations from each distribution), and the EULER inversion of the empirical transform based on the sample. Figure 2.16 shows the result of presmoothing, and Figure 2.17 shows results from exponential smoothing and the modified EULER algorithm. Table 2.2 compares the results of the three methods in terms of IAE.

Ideally we would like a more principled approach to determining the parameters of these methods, such as α, but this may not be possible given that we are seeking an approximation based on a prior belief in smoothness and unimodality. In any case,

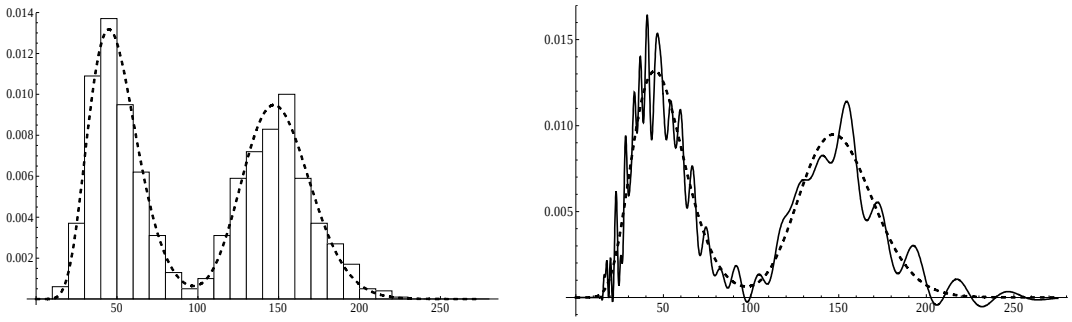


Figure 2.15: Histogram of the gamma mixture sample (left) and EULER inversion. The dashed line is the true density.

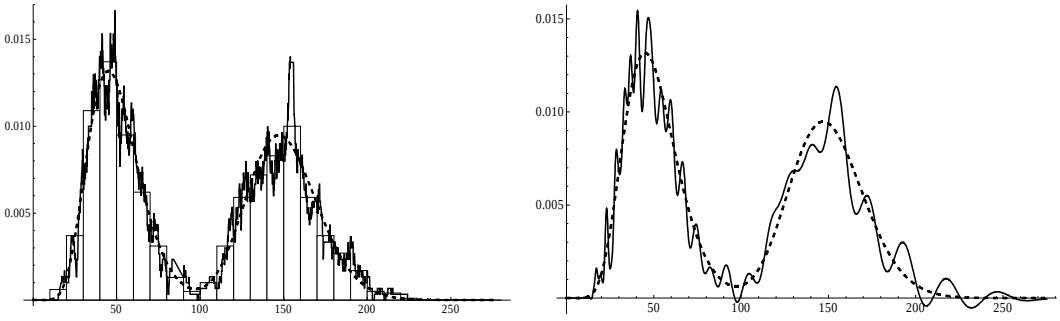


Figure 2.16: Histogram and kernel smoothing of the gamma mixture sample (left) and EULER inversion of the presmoothed sample. The dashed line is the true density.

this is a time-honored approach; the authors of a classic text on density estimation (Tapia & Thompson 1978) state: "We highly recommend the aforementioned interactive approach, where the user starts off with h_n [kernel bandwidth] values that are too large and then sequentially decreases h_n until overly noisy probability densities are obtained." "Overly noisy," of course, being in the eye of the beholder, though the condition is fairly obvious under the assumption of smoothness.

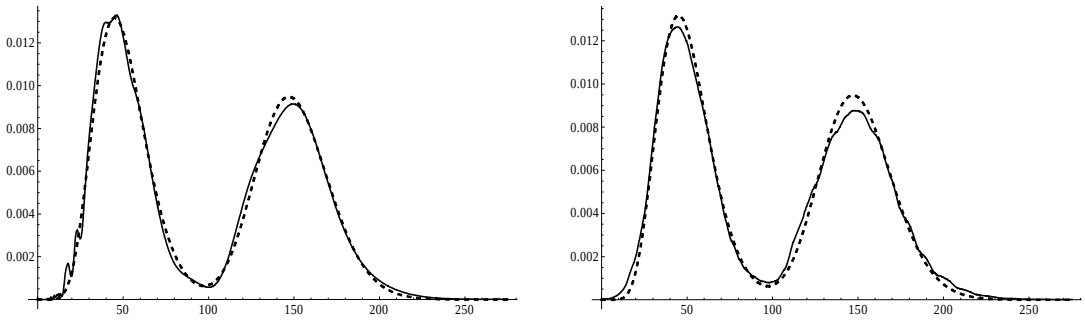


Figure 2.17: Modified EULER inversion (left) and exponential smoothing (right) of the gamma mixture sample. The dashed line is the true density.

Method	IAE
EULER	0.21659
EULER (kernel presmoothing)	0.19587
EULER (exponential smoothing)	0.07212
EULER (modified algorithm)	0.05986

Table 2.2: Integrated absolute errors (IAE) for EULER variations, gamma mixture sample.

It is possible to formalize the notion of a compromise between fit to the sample points and smoothness, using penalized maximum likelihood (Silverman 1986, Section 5.4). The penalty function is typically based on integrated derivatives of the estimated density curve (e.g., the second, which penalizes local curvature). However, this adds a nonlinear optimization problem to the computational burden, so we have not pursued it.

For approximating the distribution function, Figure 2.8 (p. 78) and the accompanying statistics illustrate the fact the numerically integrating the EULER density approximation produces a very accurate result; for that distribution, more accurate than numerical integration of the smoothed approximations. We illustrate this again

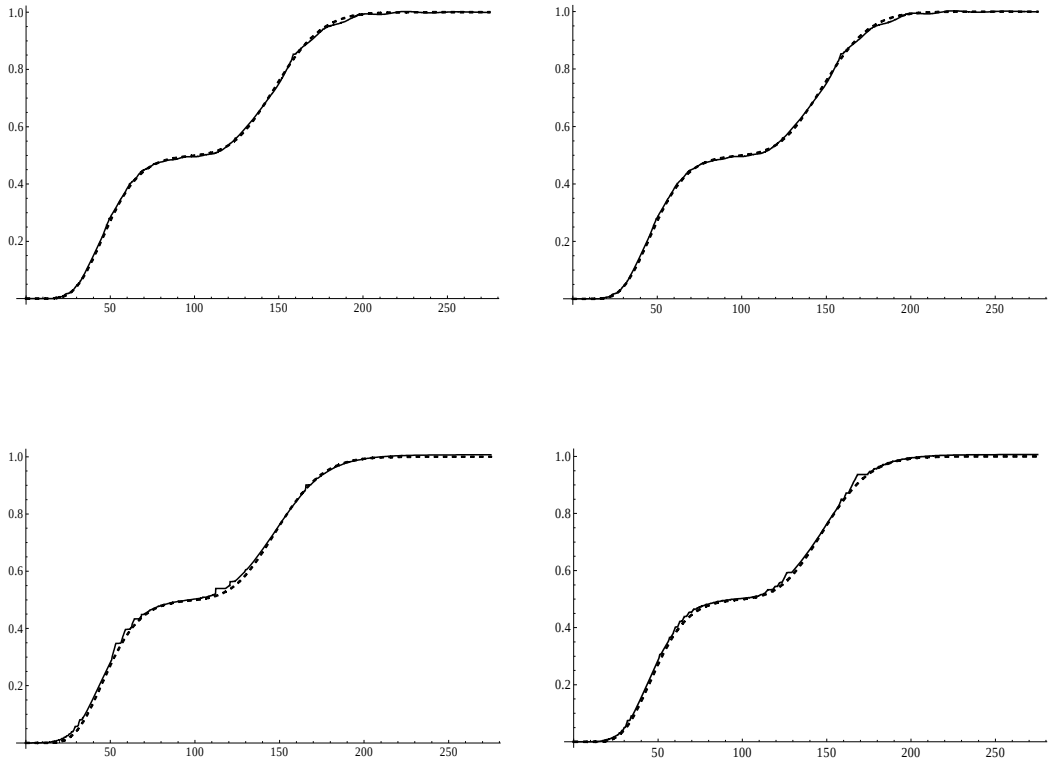


Figure 2.18: Distribution function approximations for the gamma mixture sample; clockwise from top left: EULER, kernel presmoothing, exponential smoothing, modified EULER. The dashed line is the true CDF.

with the CDF of the gamma mixture distribution. Figure 2.18 shows approximations by numerical integration of the EULER output and its smoothed versions; Table 2.3 gives the corresponding Kolmogorov-Smirnov statistics. In this case the presmoothed result is slightly better than EULER, which is better than modified EULER and exponential smoothing.

Method	K-S
EULER	0.01754
EULER (kernel presmoothing)	0.01682
EULER (exponential smoothing)	0.03742
EULER (modified algorithm)	0.03135

Table 2.3: Kolmogorov-Smirnov (K-S) statistics for gamma mixture CDF approximations.

Discussion

Based on the results shown here and other testing we have done, the best candidates for producing a smooth density estimate seem to be exponential smoothing and the modified EULER algorithm. Exponential smoothing is somewhat superior in terms of IAE and visual appearance, but modified EULER has the advantages of simplicity and reduced computational overhead. In addition, exponential smoothing and presmoothing add an additional layer of approximation, thus increasing the probability of adding error as an artifact of the smoothing process.

Computation of the distribution function by numeric integration is inherently a smoothing operation, and over a variety of distributions, integrating the output from the unmodified EULER algorithm works about as well as any of the other methods. This is to be expected, since the three smoothing methods are all, in some sense, summing the local density over intervals. For computational convenience, we use the pdf approximation plot points for approximating the CDF, so in each case both approximations are based on the same method. For the applications in Chapter 3 we will use both modified EULER and exponential smoothing.

2.2.3 Saddlepoint approximation

Given a closed form for the MGF $\mathcal{M}(s)$, the saddlepoint approximation $\hat{f}(t)$ for the density is given by

$$\hat{f}(t) = \frac{1}{\sqrt{2\pi\mathcal{K}''(\hat{s})}} \exp[\mathcal{K}(\hat{s}) - \hat{s}t] \tag{2.26}$$

where $\mathcal{K}(s) = \log[\mathcal{M}(s)]$ is the cumulant generating function (CGF) and $\hat{s}$ is the solution to the *saddlepoint equation* $\mathcal{K}'(\hat{s}) = t$. Accuracy is improved by renormalizing the density (dividing by $\int \hat{f}(t)\,dt$ so the pdf integrates to 1). Since (2.26) relies only on numerical computations based on the MGF, the EMGF $\tilde{\mathcal{M}}(s)$ may be substituted to obtain an empirical version of the saddlepoint approximation. We use the term "parametric saddlepoint approximation" to refer to the saddlepoint approximation based on an exact parametric MGF, to differentiate it from the empirical version, which is an approximation of an approximation.

Derivation of the approximation

The saddlepoint approximation was introduced to statisticians by Daniels (1954), who proved its validity using an argument based on approximating the inversion integral for the characteristic function, essentially a Laplace approximation (Copson 1965). He also presented an alternative proof based on Edgeworth expansions which we outline here, because it is more instructive in seeing how the saddlepoint method produces a smooth approximation, and in comparing it to the Fourier series approximation discussed in Section 2.2.2.

If $\phi(z)$ is the standard Gaussian density $(1/\sqrt{2\pi})e^{z^2/2}$, under certain regularity conditions the density $g(z)$ of a random variable Z with mean 0 and variance 1 can

be expanded in a convergent series

$$g(z) = c_0\phi(z) + \frac{c_1}{1!}\phi'(z) + \frac{c_2}{2!}\phi''(z) + \cdots$$

where $c_n = (-1)^n \int_{-\infty}^{\infty} H_n(z)f(z)dz$, $H_n(z)$ being the nth Hermite polynomial (Cramèr 1946, Section 17.6); this is the Gram-Charlier type A series. Where μ_i is the ith moment of the distribution of Z, the first few coefficients are $c_0 = 1$, $c_1 = c_2 = 0$, $c_3 = -\mu_3$ (up to sign, the skewness of g), and $c_4 = \mu_4 - 3$ (the excess of kurtosis over the normal distribution). After computing the derivatives the series becomes

$$g(z) = \phi(z)\left\{1 + \frac{1}{6}\mu_3(z^3 - 3z) + \frac{1}{24}\left(\mu_4 - 3\right)\left(z^4 - 6z^2 + 3\right) + \cdots\right\} \tag{2.27}$$

Since the first term, $\phi(z)$, is the approximation given by the central limit theorem, the remaining terms can be seen as corrections to the CLT based on skewness, kurtosis, and higher-order moments of $g(z)$.

Now let Y be a random variable with mean μ, variance σ^2 and density $f(y)$, and let $Z = (Y - \mu)/\sigma$. For a zero-mean RV the cumulants $\kappa_i = \mathcal{K}^{(i)}(0)$ are equal to the moments; except for the first, the cumulants of Z are the standardized cumulants of Y:

$$\rho_i = \frac{\kappa_i}{\sigma^i} = \frac{\kappa_i}{\kappa_2^{i/2}} = \frac{\mathcal{K}^{(i)}(0)}{\mathcal{K}''(0)^{i/2}}.$$

Using this, (2.27), and a standard transformation result for densities,

$$f(y) = \frac{1}{\sigma}\phi(z)\left\{1 + \frac{1}{6}\rho_3(z^3 - 3z) + \frac{1}{24}\rho_4(z^4 - 6z^2 + 3) + \cdots\right\}, \tag{2.28}$$

the *Edgeworth expansion* for $f(y)$. The standard use of the expansion is to approximate the sampling distribution of the mean, i.e., $z = (\bar{x} - \mu)/\sigma_X\sqrt{n}$ for an iid sample $\{x_1, \ldots, x_n\}$ where X_i has mean μ and variance σ_X^2, in which case a truncated Edgeworth expansion is a good asymptotic approximation under much milder conditions

than are required for the convergence of (2.27). Since the polynomials are unbounded for large z, the Edgeworth approximation is most accurate for z close to the mean.

The RV Y has MGF

$$\begin{aligned}\mathcal{M}(s) &= \int_{-\infty}^{\infty} e^{sy} f(y)dy = e^{\mathcal{K}(s)} \quad \text{so} \\ 1 &= \int_{-\infty}^{\infty} e^{sy-\mathcal{K}(s)} f(y)dy,\end{aligned}$$

thus $f(y;s) = e^{sy-\mathcal{K}(s)} f(y)$ is a density, the *exponential tilting* of $f(y)$. Let Y_s be an RV with density $f(y;s)$; it follows by direct computations that $\mathrm{E}(Y_s) = \mathcal{K}'(s)$, $\mathrm{Var}(Y_s) = \mathcal{K}''(s)$, and Y_s has the same standardized cumulants as Y.

Since $f(y) = e^{\mathcal{K}(s)-sy} f(y;s)$, an approximation for $f(y)$ can be computed from an Edgeworth expansion for $f(y;s)$. The point of this indirection is that it allows the Edgeworth expansion to be done around the mean of $f(y;s)$, where it is most accurate, for *any* value of y. This follows from the fact that $\mathcal{K}'(\hat{s}) = y$ has a solution $\hat{s}$ for y over the support of its distribution under mild conditions that are easily checked in practice (Daniels 1954, Section 6). Thus applying (2.28) to $f(y;\hat{s})$, $z = (y-\mu)/\sigma = [y - \mathcal{K}'(\hat{s})]/\sqrt{\mathcal{K}''(\hat{s})} = 0$ and

$$f(y;\hat{s}) = \frac{1}{\sqrt{2\pi\mathcal{K}''(\hat{s})}} \left\{ 1 + \frac{1}{8}\rho_4 + \cdots \right\}.$$

The first-order saddlepoint approximation is then

$$f(y) \approx \hat{f}(y) = \frac{1}{\sqrt{2\pi\mathcal{K}''(\hat{s})}} \exp[\mathcal{K}(\hat{s}) - \hat{s}y]$$

as in (2.26), where all but the first term of the Edgeworth expansion have been dropped. By including higher-order terms we get the second-order saddlepoint approximation (Daniels 1954, Equation 2.6)

$$\hat{f}(y) = \frac{1}{\sqrt{2\pi\mathcal{K}''(\hat{s})}} \exp[\mathcal{K}(\hat{s}) - \hat{s}y] \left\{ 1 + \frac{1}{8}\frac{\mathcal{K}^{(4)}(\hat{s})}{[\mathcal{K}''(\hat{s})]^2} - \frac{5}{24}\frac{[\mathcal{K}^{(3)}(\hat{s})]^2}{[\mathcal{K}''(\hat{s})]^3} \right\} \tag{2.29}$$

Higher-order approximations are also possible, but difficult because of the derivative computations; for empirical transforms, use of derivatives higher than the fourth renders the problem virtually intractable.

Details of the derivations can be found in (Daniels 1954), (Jensen 1995), or (Butler 2007); (Goutis & Casella 1999) is a good brief overview. (Cramèr 1946) has a rigorous treatment of the Gram-Charlier and Edgeworth expansions. (Reid 1988) reviews saddlepoint literature and applications, and (Huzurbazar 2005a) has many detailed examples of the saddlepoint approximation for parametric flowgraphs.

The saddlepoint density approximation

The saddlepoint method in statistics was developed to provide an asymptotically accurate approximation for the sampling density of the mean $\bar{x}$ of a sample of size n, in which case the approximation is

$$\hat{f}_n(\bar{x}) = \frac{n}{\sqrt{2\pi\mathcal{K}''(\hat{s})}} \exp(n\mathcal{K}(\hat{s}) - \hat{s}\bar{x}). \tag{2.30}$$

The error $|\hat{f}_n(\bar{x}) - f_n(\bar{x})|$ in (2.30) is $O(n^{-1})$, which is consistent with the view that it is an "improved central limit theorem"; the error in the CLT is $O(n^{-1/2})$. The saddlepoint density approximation (2.26) substitutes 1 for n in (2.30). Though the error estimate for $n = 1$ is meaningless, considerable experience validates its use to approximate $f(x)$ (Daniels 1982; Field and Ronchetti 1990; Huzurbazar 2005a; Butler 2007).

In certain cases, the saddlepoint density approximation is *exact*, as the following example shows. Suppose $X \sim \exp(\lambda)$ with density $f(x) = \lambda e^{-\lambda x}$. Then $\mathcal{M}(s) = \lambda/(\lambda - s)$, $\mathcal{K}(s) = \log\lambda - \log(\lambda - s)$, $\mathcal{K}'(s) = 1/(\lambda - s)$, and $\mathcal{K}''(s) = 1/(\lambda - s)^2$. The solution to the saddlepoint equation at x is $\hat{s} = \lambda - 1/x$, and clearly $\mathcal{K}'(\hat{s}) = x$ and

$\mathcal{K}''(\hat{s}) = x^2$. Substituting these into (2.26) and simplifying yields

$$\hat{f}(x) = \frac{e}{\sqrt{2\pi}}\lambda e^{-\lambda x}.$$

Since $e/\sqrt{2\pi}$ is independent of x, the result is exact after renormalization.

Daniels (1980) showed that the saddlepoint density approximation is exact for the Gaussian, inverse Gaussian, and gamma distributions (the exponential is a special case of the gamma), and only for those. He also conjectured (Daniels 1982) that the saddlepoint method provides a good approximation to the density for distributions that approximate the exact ones; the authors cited above provide confirmatory examples. This conjecture, which we take as a working heuristic, means that the approximation is accurate for the unimodal densities commonly encountered in reliability and survival analysis, such as gamma and Weibull. (Figure 2.19 shows an example for the Weibull density.) Nevertheless, there is no general error bound for the saddlepoint density approximation, which impacts the ability to develop confidence bands for the density, as will be discussed in Section 2.3.

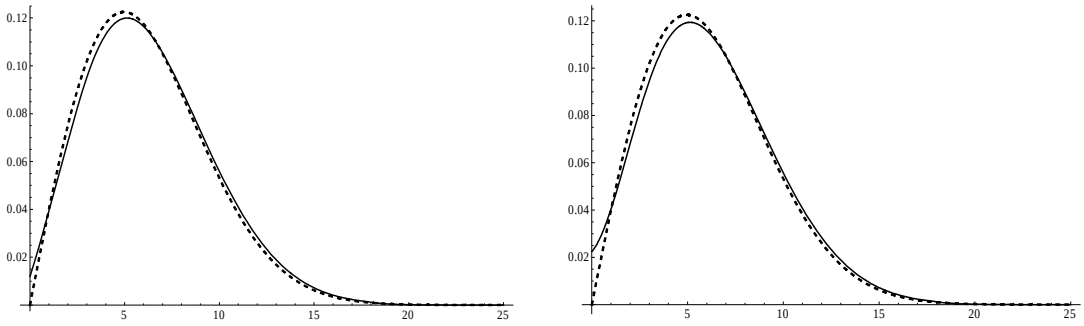


Figure 2.19: Weibull(2,7) density: parametric saddlepoint approximation (left); empirical saddlepoint approximation, n = 1,000, (right); true density is the dashed line.

To illustrate the potential problem, consider the density

$$f(t) = \frac{\lambda\pi}{2 - 2e^{-\lambda\pi}} \left(e^{-\lambda\pi t} + e^{\lambda\pi(t-1)}\right) I_{[0,1]}(t).$$

This resembles a beta density, but has a more tractable MGF. The density for $\lambda = 2.5$ and its parametric saddlepoint approximation are plotted in Figure 2.20; the IAE of the saddlepoint approximation is 0.7336, and the approximation has little use other than giving an indication of the general shape of the density. Section 7.2 of (Field and Ronchetti 1990) has other examples like this.

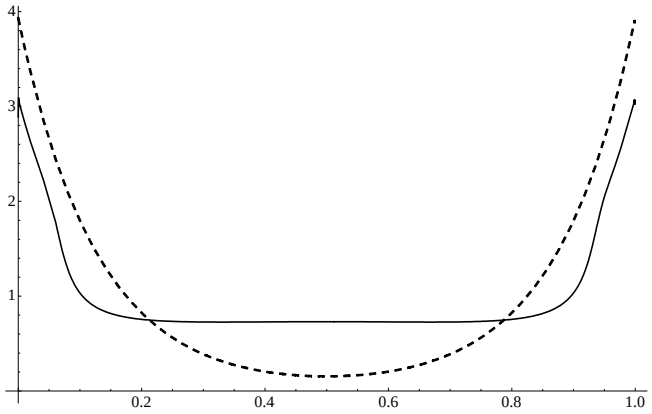


Figure 2.20: Saddlepoint density approximation; the true density is the dashed line.

As a more practical example, consider the gamma mixture that was used as an example in the previous section (p. 86). This is the sort of distribution that appears in reliability and survival analysis when there are competing risks from multiple failure causes or diseases; a real example will be described in detail in Section 3.2. Figure 2.21 shows first and second-order parametric saddlepoint density approximations; the first-order, in particular, is quite misleading as to the shape of the density. (Note the formation of a mode at $t = 100$, the mean of the distribution; this is a consequence of the central limit basis for the approximation.) Comparison

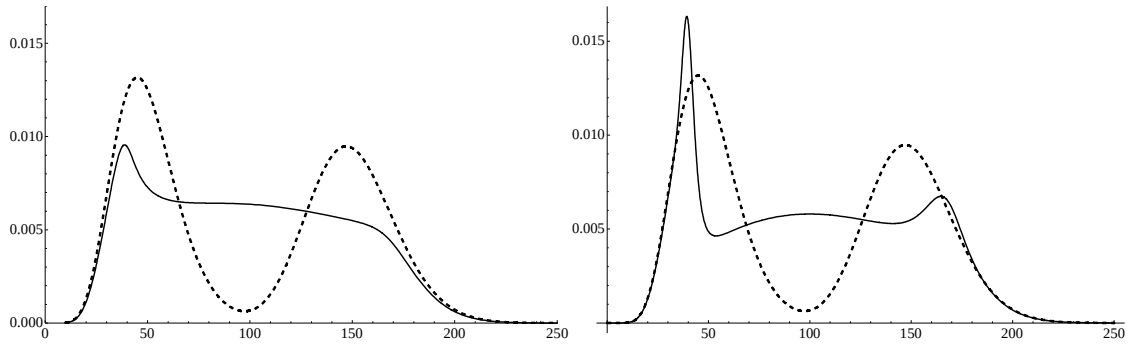


Figure 2.21: First-order (left) and second-order saddlepoint approximations of the gamma mixture density; the true density is the dashed line.

of this to Figure 2.17 (p. 89) shows that the empirical EULER inversion is superior to the parametric saddlepoint approximation, and *a fortiori* will be superior to the empirical saddlepoint approximation.

How large can the error be in the saddlepoint density approximation? The following theorem gives a pessimistic answer:

Theorem 2.2.1 (Possible error in the saddlepoint density approximation) There exists $f(t)$, a continuous pdf with first-order saddlepoint density approximation $\hat{f}(t)$, and $\mathcal{I} = (t_0 - \epsilon, t_0 + \epsilon)$, an interval in the support of f with finite measure, such that $|f(t) - \hat{f}(t)|$ is greater than any positive number M for all $t \in \mathcal{I}$.

Proof: We prove the result by demonstrating the existence of such a density function. Let $X \sim$ exponential(1) and $Y \sim$ gamma$(1/\delta, \delta)$ with pdfs

$$f_X(t) = e^{-t}, \qquad f_Y(t) = \frac{1}{\Gamma(1/\delta)\delta^{1/\delta}} t^{1/\delta - 1} e^{-t/\delta} \tag{2.31}$$

where $\delta > 0$ is a small undetermined constant. Let Z be a mixture of X and Y with pdf $f_Z(t) = .5f_X(t) + .5f_Y(t)$. Figure 2.22 shows $f_Z(t)$ and its saddlepoint

approxmation for $\delta = .001$. We will will show that as $\delta \to 0$ the saddlepoint approximation at the mean remains less then 1, while the actual density (the sharp peak) is unbounded.

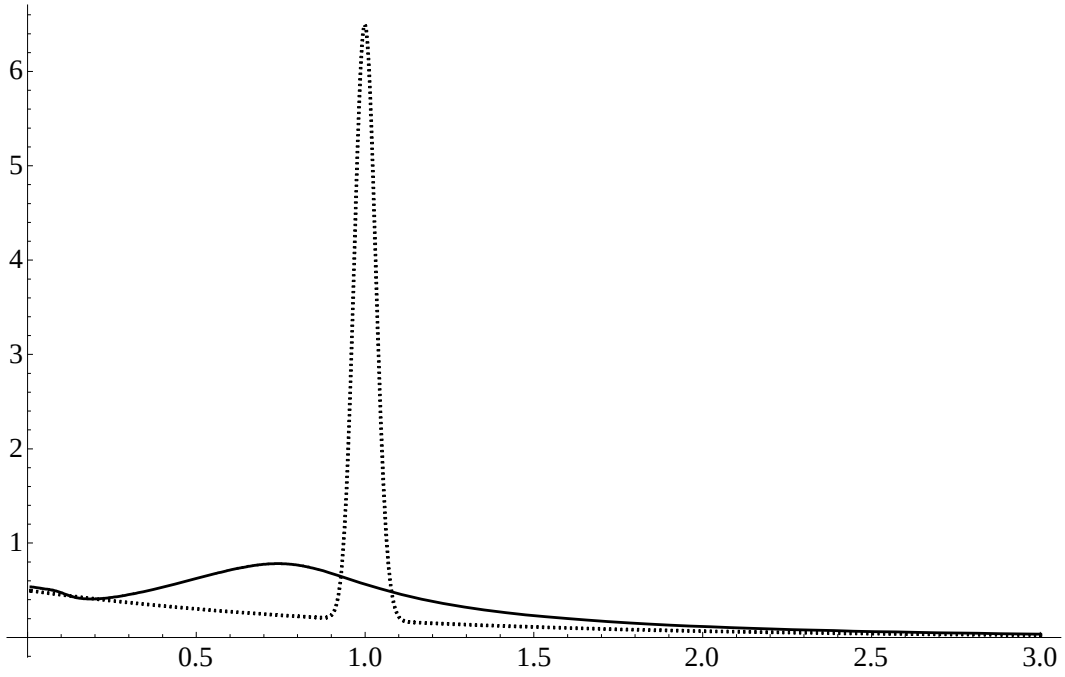


Figure 2.22: First-order saddlepoint approximation to an exponential-gamma mixture density; the true density is the dotted line.

The MGFs for X and Y and their derivatives are

$$\begin{aligned}
\mathcal{M}_X(s) &= \frac{1}{1-s}\,, \quad \mathcal{M}'_X(s) = \frac{1}{(1-s)^2}\,, \quad \mathcal{M}''_X(s) = \frac{2}{(1-s)^3}\,, \\
\mathcal{M}_Y(s) &= \frac{1}{(1-\delta s)^{1/\delta}}\,, \quad \mathcal{M}'_Y(s) = \frac{1}{(1-\delta s)^{1/\delta+1}}\,, \quad \mathcal{M}''_Y(s) = \frac{1+\delta}{(1-\delta s)^{1/\delta+2}}.
\end{aligned}$$

The CGF for Z and its derivatives are (dropping the argument s for brevity)

$$\begin{aligned}\mathcal{K}_Z &= \log(.5\,\mathcal{M}_X + .5\,\mathcal{M}_Y)\,, \qquad \mathcal{K}'_Z = \frac{\mathcal{M}'_X + \mathcal{M}'_Y}{\mathcal{M}_X + \mathcal{M}_Y}\,,\\ \mathcal{K}''_Z &= \frac{(\mathcal{M}_X + \mathcal{M}_Y)(\mathcal{M}''_X + \mathcal{M}''_Y) - (\mathcal{M}'_X + \mathcal{M}'_Y)^2}{(\mathcal{M}_X + \mathcal{M}_Y)^2}.\end{aligned}$$

The common mean of X, Y, and Z is 1, and we now consider the values of $f_Z(t)$ and its saddlepoint approximation at the mean. $\mathcal{M}_X(0) = \mathcal{M}_Y(0) = \mathcal{M}'_X(0) = \mathcal{M}'_Y(0) = 1$, so by inspection the solution of the saddlepoint equation $\mathcal{K}'_Z(\hat{s}) = 1$ is $\hat{s} = 0$. By simple computations

$$\mathcal{M}''_X(0) = 2, \quad \mathcal{M}''_Y(0) = 1 + \delta, \quad \mathcal{K}''_Z(0) = \frac{2+\delta}{4}.$$

Substituting these values in the first-order saddlepoint formula (2.26), the density approximation is

$$\begin{aligned}\hat{f}_Z(1) &= \frac{1}{\sqrt{2\pi\mathcal{K}''_Z(0)}}\, e^{\mathcal{K}_Z(0)}\\ &= \frac{e}{\sqrt{\pi(1+\delta/2)}}\end{aligned}$$

This increases continuously as $\delta \to 0$, so its supremum is

$$\lim_{\delta\to 0} \hat{f}_Z(1,\delta) = \frac{e}{\sqrt{\pi}} \approx .8652.$$

From the definition of Z, $f_Z(t) \geq .5\, f_Y(t)$, so if $f_Y(1)$ is unbounded as $\delta \to 0$, so is $f_Z(1)$. Since f_Y is continuous in δ, to determine its limit at $t = 1$ when $\delta \to 0$ it suffices to consider only values of δ such that $1/\delta = k \in \mathbb{N}$, thus

$$\begin{aligned}\lim_{\delta\to 0} f_Y(1,\delta) &= \lim_{\delta\to 0} \frac{1}{\Gamma(1/\delta)\delta^{1/\delta}} e^{-1/\delta}\\ &= \lim_{k\to\infty} \frac{1}{\Gamma(k)(1/k)^k} e^{-k}\\ &= \lim_{k\to\infty} \frac{1}{(k-1)!(1/k)^k} e^{-k}.\end{aligned}$$

Now using Stirling's approximation, $(k-1)! \sim (k-1)^{k-1/2}\sqrt{2\pi}\, e^{-(k-1)}$,

$$\begin{aligned}
\lim_{\delta\to 0} f_Y(1,\delta) &= \lim_{k\to\infty} \frac{e^{-k}}{(k-1)^{k-1/2}\sqrt{2\pi}\, e^{-(k-1)}(1/k)^k} \\
&= \lim_{k\to\infty} \frac{e^{-1}}{(k-1)^{-1/2}(1-1/k)^k\sqrt{2\pi}} \\
&= \lim_{k\to\infty} \sqrt{\frac{k-1}{2\pi}} \quad \text{using the fact that } (1-1/k)^k \to e^{-1} \\
&= \infty.
\end{aligned}$$

So $f_Y(1)$, and therefore $f_Z(1)$, is unbounded as $\delta \to 0$. Since $\hat{f}_Z(1)$ is bounded by 1, given any $M > 0$, there is a value δ_0 such that if $\delta < \delta_0$, then $|\hat{f}_Z(1) - f_Z(1)| > M$. By the continuity of f_Z, this condition can be satisfied for t close enough to 1, i.e., $|\hat{f}_Z(1 \pm \epsilon) - f_Z(1 \pm \epsilon)| > M$ for small ϵ. Thus for any $M > 0$ there is an interval $\mathcal{I} = (1-\epsilon, 1+\epsilon)$ within which the error in the saddlepoint density approximation exceeds M. ∎

The empirical saddlepoint approximation

To compute $\tilde{f}(t)$, the empirical saddlepoint density, we follow the same procedure using empirical versions of $\mathcal{K}(s)$ and its derivatives. In the simplest case, given sample data $t_1, \ldots, t_n$ from a single transition $j \to k$, we define the empirical CGF in the obvious way as (Davison & Hinkley 1988):

$$\tilde{\mathcal{K}}(s) = \log[\tilde{\mathcal{M}}(s)] = \log\left[\frac{1}{n}\sum_{i=1}^{n} e^{st_i}\right]$$

from which the derivatives are calculated as

$$\begin{aligned}
\tilde{\mathcal{K}}'(s) &= \frac{\sum_i t_i e^{st_i}}{\sum_i e^{st_i}} \qquad (2.32) \\
\tilde{\mathcal{K}}''(s) &= \frac{[\sum_i e^{st_i}]\,[\sum_i t_i^2 e^{st_i}] - [\sum_i t_i e^{st_i}]^2}{[\sum_i e^{st_i}]^2}.
\end{aligned}$$

From the convergence of $\tilde{\mathcal{M}}(s)$ to $\mathcal{M}(s)$ (see Section 2.1.3) and from standard theorems of analysis, it follows that the same convergence holds for derivatives $\tilde{\mathcal{M}}^{(k)}(s)$, for $\tilde{\mathcal{K}}(s)$ and its derivatives, and for the empirical saddlepoint approximation itself; i.e., if $\tilde{f}_n(t)$ is the empirical saddlepoint approximation based on samples with a minimum size of n, and $\hat{f}(t)$ is the parametric saddlepoint approximation, $\tilde{f}_n(t) \to \hat{f}(t)$ almost surely as $n \to \infty$. Proofs can be found in (Feuerverger 1989).

Unlike the parametric case, here it cannot be guaranteed that the saddlepoint equation has a solution over the entire support of the true density. The following theorem, which is alluded to but not proved in (Davison & Hinkley 1988), makes this more precise.

Theorem 2.2.2 (Empirical saddlepoint solution) Suppose $\tilde{\mathcal{K}}(s)$ is an empirical CGF with basis sample $t_1, \ldots, t_n$; assume that the t_i do not all have the same value. Then the empirical saddlepoint equation $\tilde{\mathcal{K}}'(\hat{s}) = t$ has a unique real solution $\hat{s}$ if and only if $t_{(1)} < t < t_{(n)}$, where the $t_{(i)}$ are the order statistics of the sample.

Proof: From (2.32), a solution exists iff

$$\sum_{i=1}^{n} (t_i - t) e^{s t_i} = 0 \quad \text{for some } s \in \mathbb{R}. \tag{2.33}$$

If $t \leq t_{(1)}$ then every term on the left in (2.33) is positive or zero, with at least one nonzero (since the t_i are not identically equal); if $t \geq t_{(n)}$ then every term is negative or zero, with at least one nonzero; in either case there is no real solution. So it remains to prove that there is a unique real solution for $t_{(1)} < t < t_{(n)}$.

Multiply both sides of (2.33) by e^{-t}; then a solution exists iff

$$\sum_{i=1}^{n} (t_i - t) e^{s(t_i - t)} = 0 \quad \text{for some } s \in \mathbb{R}. \tag{2.34}$$

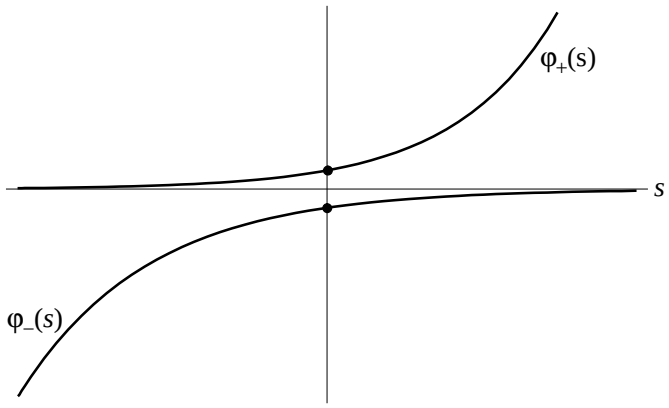


Figure 2.23: Solution of the empirical saddlepoint equation

Let $\varphi_+(s) = \sum_{\{i:\, t_i - t \geq 0\}} (t_i - t) e^{s(t_i - t)}$, $\varphi_-(s) = \sum_{\{i:\, t_i - t < 0\}} (t_i - t) e^{s(t_i - t)}$. Figure 2.23 shows, for real s, the general shape of $\varphi_+(s)$, which is convex and always positive, and $\varphi_-(s)$, which is concave and always negative. Both functions are continuous. By (2.34) a solution $\hat{s}$ exists iff $|\varphi_+(\hat{s})| = |\varphi_-(\hat{s})|$. Consider three cases: if $|\varphi_+(0)| = |\varphi_-(0)|$, then $\hat{s} = 0$; if $|\varphi_+(0)| > |\varphi_-(0)|$, then since $|\varphi_-(s)|$ is increasing to the left of the origin and $|\varphi_+(s)|$ is decreasing to the left of the origin, by continuity there must be a point $\hat{s} < 0$ where they are equal; similarly, if $|\varphi_+(0)| < |\varphi_-(0)|$, there must be a point $\hat{s} > 0$ where they are equal, so in any case a solution exists.

To show uniqueness, consider (2.34) and let $\varphi(s) = \sum_i (t_i - t) e^{s(t_i - t)}$. By assumption $t_i - t$ is not identically 0, so $\varphi'(s) = \sum_i (t_i - t)^2 e^{s(t_i - t)} > 0$ for $s \in \mathbb{R}$; thus $\varphi(s)$ is a strictly increasing function and cannot have more than one zero on $\mathbb{R}$. ∎

As a corollary of the first passage sample theorem (Theorem 2.1.1, p. 57), a similar restriction of the domain applies in the general case where $\tilde{\mathcal{M}}(s)$ is an algebraic combination of transition EMGFs $\tilde{\mathcal{M}}_{ij}(s)$. In that case $t_{(1)}$ will be the smallest possible first passage time computed from the basis samples for the EMGFs, and $t_{(n)}$ will be the largest possible first passage time. For flowgraphs with loops, $t_{(n)}$ is

effectively infinite (see p. 44, part (5) of Lemma 2.1.2).

In practice, determining the exact limits for solving the empirical saddlepoint equation is not generally feasible, so we determine the limits numerically. This is illustrated in Figure 2.24. $\mathcal{K}'(s)$, the exact derivative of the CGF for a gamma(2,5) distribution, is asymptotic to 0 at $-\infty$ and has a simple pole at .2, so $\hat{s}$ exists for any value of t. $\tilde{\mathcal{K}}'(s)$ is asymptotic to .212204 (the smallest value in this particular sample) on the left and to 41.0824 on the right (the largest value in the sample), so $\hat{s}$ exists only for $t \in (.212204, 41.0824)$. This sort of graphical determination can be done for any empirical flowgraph problem, and is used to set the root-finding limits in the saddlepoint approximation code (see Appendix A.4). The output from the saddlepoint approximation code is an interpolating function (see Appendix A.1.2) which extrapolates linearly to compute values of the approximation outside the sample limits.

One problem that might be introduced by linear extrapolation would be missing a mode of the density; however, by definition of a mode, it is likely that there would be sample points on both sides of it, so the problem should rarely occur. More generally, the lack of information beyond the sample limits makes the approximation less accurate in the tails, which will be seen in the examples in Chapter 3.

Saddlepoint CDF estimation

There is also a saddlepoint formula based on the CGF for directly approximating points of the distribution function, due to Lugannani and Rice (1980) . This method is not claimed to be more accurate (*ibid.*, p. 480), but is more convenient if only a few percentiles of the distribution are needed. The Lugannani and Rice method is not immune to the sort of accuracy problems described for the density approximation (Booth & Wood 1995), and exhibits certain numerical instabilities (Huzurbazar &

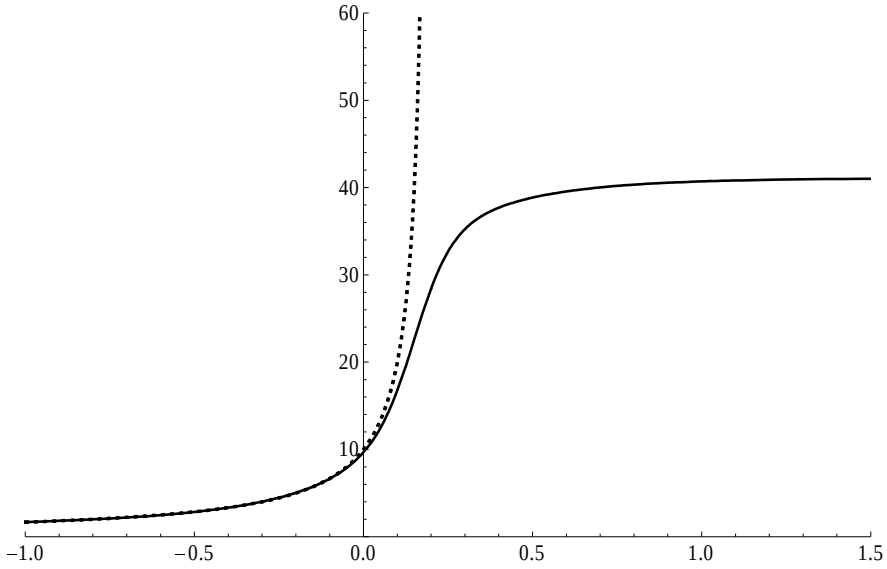


Figure 2.24: $\tilde{\mathcal{K}}'(s)$ (solid line) and $\mathcal{K}'(s)$ (dotted line) for real values of s; gamma(2,5), sample size 1,000.

Huzurbazar 1999).

Our preferred alternative is to numerically integrate the saddlepoint density approximation, since it is more convenient given that we want to plot the entire CDF, compute the hazard function, and compute various statistics such as the Kolmogorov-Smirnov distance. Numerical integration is also consistent with the computation of the EULER CDF approximation. Figure 2.25 shows CDF approximations for the Weibull(2,7) distribution, computed by numerically integrating the saddlepoint density approximations (shown in Figure 2.19.) The empirical result is actually better then the parametric result in the left tail—this is because the Weibull MGF is complicated and numerically unstable for negative s, causing a problem for values of t near the origin. Table 2.4 shows error statistics for the approximations.

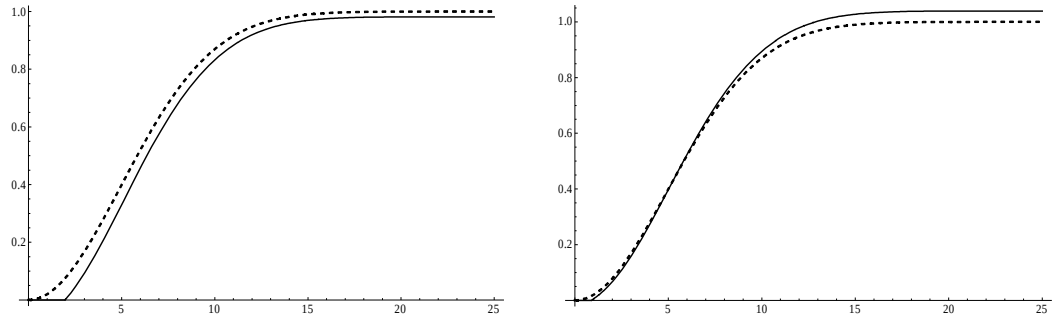


Figure 2.25: Weibull(2,7) CDF—parametric saddlepoint approximation, (left); empirical saddlepoint approximation, $n = 1{,}000$; true CDF is the dashed line.

Method	IAE (pdf)	K-S
Parametric saddlepoint	0.04713	0.07565
Empirical saddlepoint	0.05253	0.03914

Table 2.4: Integrated absolute errors (IAE) for pdf and Kolmogorov-Smirnov (K-S) statistics for Weibull(2,7) approximations.

Discussion

The saddlepoint density approximation is smooth, and remarkably accurate for a certain class of densities. For others, particularly mixture densities resulting from competing risks, it can have serious errors because it is pulled towards a central limit approximation. EULER is accurate over a broader class or densities, but only in a certain average sense, due to its approximation using an oscillatory integrand. In some cases, but not all, the oscillations can be smoothed out for an accurate result. By using both methods we usually achieve a result in which we can have some confidence, at least to the extent of validating (or falsifying) proposed parametric models.

2.2.4 Potential Bayesian approaches

In the literature on nonparametric density or curve estimation one often finds statements like "Methods of nonparametric curve estimation allow one to analyze . . . data at hand without any prior opinion about the data" (Efromovich 1999, p. 1); yet in the same text we find this: "Thus, the last step is to remove bumps that *we believe* are due solely to the oscillatory approximation of the flat parts of an underlying density" (*ibid.*, p. 64, emphasis added). Even the innocent histogram embodies assumptions about the "graininess" of the distribution (e.g., whether it is multimodal) in the choice of bin sizes and placements; kernel smoothing uses the assumption of a continuous density, as well as assumptions about the modality of the distribution which influence the choice of kernel bandwidth. And of course, in the present work we are explicitly basing our inversion methods on an assumption of continuous densities. Arguably the only completely "frequentist" nonparametric density estimate is the nonparametric maximum likelihood estimate (NPMLE), which for a sample of size n is a discrete distribution placing probability $1/n$ at each sample point (Owen 2001). Any other method is at least philosophically Bayesian, in that it uses some prior belief about properties of the distribution.

Philosophy aside, there are technical Bayesian approaches to density estimation which are nonparametric or semiparametric, in the sense of using priors over families of distributions that are extremely large—e.g., all continuous distributions, or all distributions constructed piecewise or as mixtures from members of a given family. (Hjort 1996) and (Ghosh & Ramamoorthi 2003, Chapter 5) provide overviews of such "nonparametric Bayesian" density estimation.

In the parametric flowgraph context, parametric Bayesian methods have been applied by specifying a prior for each holding time distribution and transition probability in the model, then using the standard Bayesian machinery to compute a pos-

terior distribution for the model parameters. In general, numerical methods must be used to sample from the posterior, compute the saddlepoint approximation $\hat{f}(t|\,\boldsymbol{\theta}_i)$ at a point t for each sampled parameter vector $\boldsymbol{\theta}_i$, then average over the $\hat{f}(t|\,\boldsymbol{\theta}_i)$ to get an approximation to the posterior predictive density at t. See (Huzurbazar 2005c; Huzurbazar 2005a, Chapter 5) for details and examples. In the context of a semiparametric or mixed flowgraph model (see Section 2.4), these standard Bayesian methods may be applied to the transitions that are modeled parametrically.

In a fully nonparametric context, there are several ways we might proceed. Using a Bayesian nonparametric method we could obtain a posterior predictive density for each holding time distribution, sample it, and compute the empirical transforms from the posterior samples. Or, we could work through the nonparametric flowgraph method in the standard way, then sample from the first passage density approximation and apply a Bayesian nonparametric density estimation method to the sample. The latter method might be appropriate if there were a reasonable prior belief in some property of the first passage distribution, without prior beliefs in the details of individual transitions. A third possibility, described below, is to use priors on the coefficients of a series expansion approximating the first passage density.

As an example, (Escobar & West 1995) describes density estimation using a Dirichlet mixture of normals. In a simple variant, the sample data points are assumed to be independently distributed as $X_i \sim N(\mu_i, \sigma^2)$ with common variance σ^2 and μ_i drawn from a distribution F, with the prior distribution on F a Dirichlet process (Ferguson 1973). The use of the normal distribution introduces a parametric element, but note that if the μ_i are drawn from a Dirichlet distribution with mass concentrated at the observed sample points and the mixing parameter is identically $1/n$, this yields essentially the density approximation by a mixture of Gaussians that was described earlier in this section (p. 72). Thus Bayesian density estimation with Dirichlet mixtures can be seen as a flexible generalization of kernel smoothing.

A conceptually related idea is the mixture of finite Pólya trees (Hanson 2006), which generalizes an underlying parametric density by recursively dividing its support into dyadic subintervals and scaling the height of the density in each subinterval by an additional parameter representing the probability that it contains the value of the random variable; this recursive division can be carried out an arbitrary number of times. See (Christensen *et al.* 2008) for computational details and examples. Neither this method nor the Dirichlet mixture is restricted to using the Gaussian as the base distribution; in the context of reliability and survival analysis a distribution on nonnegative random variables such as the gamma, inverse Gaussian or Weibull might be more appropriate.

Since both the Fourier series and saddlepoint inversions are based on expansions in orthogonal functions (the latter via the Edgeworth series), one might consider, as suggested by Hjort (1996, Section 4), placing a prior on the series coefficients. To do so in a principled way the coefficients should be interpretable as properties about which one could have prior beliefs. This seems difficult in the case of the Fourier series, but is possible for the saddlepoint approximation, since the coefficients in the Edgeworth series can be interpreted as corrections to a normal approximation based on moments of the sampled distribution. Using a higher-order saddlepoint approximation would allow exploitation of prior beliefs about moments, but also requires computation of the higher-order derivatives of the empirical CGF, which can be quite complicated.

Bayesian methods of the sort just described are promising because in applications one can often elicit quantitative prior beliefs regarding holding time distributions; these may be based on scientific theory or prior experience with similar systems. The added computational burden associated with these methods has so far deterred us from using them. However, with anticipated improvements in the inversion algorithms for empirical transforms they may become feasible, so we mention them in

the interest of completeness and pointing out directions for future research.

2.3 Confidence bounds

Having an estimate $\tilde{g}_{ij}$ of the first-passage density g_{ij}, or by numerical integration an estimate $\tilde{G}_{ij}$ of the first-passage distribution function G_{ij}, an obvious further question is whether we can obtain confidence bounds of various sorts. E.g., for a particular τ_0 and $0 < \alpha < 1$, can we find d_l and d_u such that

$$\mathrm{P}\{\tilde{G}_{ij}(\tau_0) - d_l \leq G_{ij}(\tau_0) \leq \tilde{G}_{ij}(\tau_0) + d_u\} \geq 1 - \alpha, \tag{2.35}$$

or can we construct a confidence band, say, around the estimate of G_{ij} so that (2.35) holds pointwise for all τ in the support of G_{ij}? (Statements like this are to be interpreted in a frequentist sense— $G_{ij}(\tau_0)$ is taken to be an unknown constant, $\tilde{G}_{ij}(\tau_0)$ to be random, dependent on the basis samples, and (2.35) asserts that the probability of $G_{ij}(\tau_0)$ lying in the random interval $[\tilde{G}_{ij}(\tau_0) - d_l, \tilde{G}_{ij}(\tau_0) + d_u]$ is at least $1 - \alpha$.)

Unfortunately the answer to these questions is largely negative, due to the methods we use for estimating the density (and subsequently the CDF). This is somewhat surprising, since if we had an iid sample $\tau_1, \tau_2, \ldots, \tau_n$ of first passage times, the standard nonparametric estimate of G_{ij} would be the empirical distribution function $\hat{G}_{ij,\,n}(\tau) = n^{-1} \sum_{m=1}^{n} I_{[0,\tau]}(\tau_m)$, for which large sample properties such as confidence bands are well known; see, e.g., (Serfling 1980, Chapter 2). However, since our distribution estimates are not constructed directly from an iid sample, the usual confidence bands or intervals based on properties of the EDF are not applicable.

Even with an iid sample of first passage times, confidence bands for the pdf are more problematic, depending in general on properties of the true density (Silverman

1986). This leads into an explanation of problems with the methods we use for estimating the first passage density. Silverman (1986, p. 37) says about the general problem that "It is characteristic of almost all density estimation methods that the estimate is of the form

smoothed version of true density + random error

where the 'smoothed version of the true density' depends deterministically on the precise choice of parameters in the method, but not directly on the sample size." I.e., understanding the small or large sample distribution for the random error due to sampling is not sufficient for determining confidence bands; it is also necessary to quantify the "structural bias" introduced by the estimation method.

In the case of the EDF, there is no bias— $\mathrm{E}[\hat{G}_{ij,\,n}(\tau)] = G_{ij}(\tau)$, and the error is entirely due to sampling. For estimation of the sort we do, the situation is more complex, as shown by consideration of an identity based on Equation 3.6 in (Feuerverger 1989), which is similar to Silverman's above. Let $f(x)$ be the pdf of a random variable X, $\hat{f}_{app}(x)$ either the saddlepoint or EULER approximation based the exact parametric transform of f, and $\hat{f}_{emp}(x)$ the empirical version of the approximation based on a sample $x_1, \ldots, x_n$. Then the error in the empirical estimate at x is

$$\hat{f}_{emp}(x) - f(x) \;=\; \{\hat{f}_{emp}(x) - \hat{f}_{app}(x)\} + \{\hat{f}_{app}(x) - f(x)\}. \tag{2.36}$$

Both bracketed terms on the right-hand side are problematic. Since the empirical transform is a strongly consistent estimator, the first term includes a randomization error whose large sample properties can be found in the literature cited in Section 2.1.3. However, if f is the first passage density for a complex flowgraph, the empirical transform is given by a Mason's rule expression; while still being a strongly consistent estimator (Theorem 2.1.2), it is a rational function of RVs, whose large sample properties are, in practice, intractable (Springer 1979). The second term on the right-hand side of (2.36) is even more problematic. As shown in Section 2.2.3, the

saddlepoint density estimate does not have an error bound. Approximate bounds for the EULER approximation exist only for transforms of distributions that are absolutely continuous or lattice (Abate & Whitt 1992; Strawderman 2004). These problems preclude developing a distribution or confidence bounds for the left-hand side of (2.36).

Davison and Hinkley (1997, p. 421) make a similar point regarding the empirical saddlepoint approximation for the sampling density of the mean, concluding "There is no exact general nonparametric confidence limit procedure for the mean," because the parametric saddlepoint approximation for the sampling distribution of the mean is used in place of the true distribution, thus adding the error in the second term in the right-hand side of (2.36). *A fortiori*, their conclusion applies to the saddlepoint density approximation. As mentioned in Section 2.2.3, the same conclusion can be drawn regarding the Lugannani and Rice (1980) saddlepoint approximation for the distribution function.

Ouhbi and Limnios (2001) showed that a nonparametric solution for the renewal equation giving system reliability (equivalent to the first passage distribution to a failing state), is asymptotically normal. However, the variance depends on the unknown true value of $\mathbf{Q}$, so it cannot be used directly to compute confidence bands. One might substitute an empirical estimate of $\mathbf{Q}$, or use the reliability estimator to compute bootstrap confidence bands; but these would require some numerical technique with a known error bound for solving the empirical version of the renewal equation. We are not aware of such a method, so this runs into the problems described above (cf. Section 2.5, p. 118)

It is possible, however, to derive confidence intervals for quantities based on moments of the first passage time distribution, since for $\mathcal{M}$ the MGF, $\mathcal{M}'(0)$, $\mathcal{M}''(0), \ldots,$ $\mathcal{M}^{(r)}(0), \ldots$ give the (noncentral) moments $\mathrm{E}(\mathcal{T})$, $\mathrm{E}(\mathcal{T}^2)$, $\ldots$, $\mathrm{E}(\mathcal{T}^r)$. From Theorem 2.1.2 the EMGF is a strongly consistent pointwise estimator for val-

ues of the true MGF, so approximate moments can be gotten by substituting the empirical MGF, i.e., $\tilde{\mathcal{M}}^{(r)}(0) \approx \mathcal{M}^{(r)}(0)$. (We assume here that the moments of the true distribution exist; otherwise the approximate moments will be unbounded rather than converging to a finite value.) As discussed above, we have no distributional results for empirical transforms given by a Mason's rule expression, but we can derive a bootstrap approximation for confidence intervals around the moments.

Taking the mean as an example, and assuming $G(\tau)$ is the first passage distribution and $\bar{\tau}$ is the average of a sample $\tau_1, \ldots, \tau_n$, a $(1-\alpha)100\%$ CI is an interval $(L, U) = (\bar{\tau} - d_l, \bar{\tau} + d_u)$ such that $d_l, d_u > 0$ and

$$\mathrm{P}\{\mathrm{E}_G(\mathcal{T}) \in (L, U)\} = \mathrm{P}\{\mathcal{M}'_G(0) \in (L, U)\} \geq 1 - \alpha$$

where d_l and d_u are selected based on the sampling distribution of the mean of G. If the true G is unknown, the EDF $\hat{G}_n$ can be used as a plug-in estimate to construct d_l^* and d_u^*, leading to an asymptotically correct CI $(L^*, U^*) = (\bar{\tau} - d_l^*, \bar{\tau} + d_u^*)$.

In cases where the exact distribution needed for deriving the CI is difficult or impossible to compute, conventional bootstrapping approximates (L^*, U^*) by repeatedly drawing samples of size n (with replacement) from $\{\tau_1, \ldots, \tau_n\}$, computing $\hat{\mathcal{M}}'_{\hat{G}_n}(0) \approx \mathcal{M}'_G(0) = \mathrm{E}_G(\mathcal{T})$, and determining the (L^*, U^*) that covers $(1-\alpha)100\%$ of the values of $\hat{\mathcal{M}}'_{\hat{G}_n}(0)$. Asymptotically (as $n \to \infty$), this procedure converges to the exact CI (Davison & Hinkley 1997).

Here we do not have a sample from G, but since the empirical MGF $\tilde{\mathcal{M}}'_{\hat{G}_n}(0)$ derived using Mason's rule is a consistent estimate of $\mathcal{M}'_G(0)$, it can be used in the same way to derive an asymptotically correct CI. The bootstrapping is done by resampling each of the holding time samples and recomputing the EMGF $\tilde{\mathcal{M}}'_{\hat{G}_n}(0)$. The same procedure can be used to obtain bootstrap CIs for any moment $\mathrm{E}(\mathcal{T}^r) = \mathcal{M}_G^{(r)}(0)$. Typically the mean and variance are the quantities of interest.

Though this bootstrap CI is asymptotically (in the size of the basis samples)

correct, the speed of convergence depends on the particular problem; in addition, the particular samples we happen to have will influence the accuracy of the result. For sample sizes we can reasonably expect to encounter, the actual coverage of the bootstrap CI will usually be less than the nominal coverage. For example, in the repairable system example (Section 3.1), we use a 99% bootstrap CI to obtain actual coverage of about 95%.

If we hypothesize, based on prior knowledge, that the first passage distribution is a particular one of the types commonly encountered in reliability and survival analysis, estimates and confidence intervals for the first few moments can be used to develop distribution estimates by the method of moments (MOM). Though generally not as good as maximum likelihood estimates, unbiased and reasonably efficient MOM techniques exist for distributions such as the exponential (where the MOME is identical to the MLE), gamma (Wiens *et al.* 2003), Weibull (Singh *et al.* 1990), and inverse Gaussian (Padgett & Wei 1979). By computing MOM estimates of the pdf or CDF using moment values at the limits of the confidence intervals described above, approximate confidence bands can be constructed.

The general paucity of results for confidence bounds is a consequence of the several layers of estimation and approximation between the basis samples and a first passage density or distribution function; each layer has potential randomization and numerical errors. Currently, reasonable confidence bounds for the first passage pdf or CDF appear to be intractable. This is unfortunate, since in practical applications having such bounds is very desirable. We hope to improve on this in future research.

2.4 Mixed or semiparametric flowgraph models

Mason's rule operates on transforms, and nonparametric analysis can be viewed as substituting transforms with respect to EDFs (based on observed samples) for transforms with respect to parametric distributions. Since the algebra is indifferent to the type of transform it operates on, there is no reason why we cannot "mix and match" parametric and empirical transforms. This mixed or semiparametric analysis is potentially valuable in situations where, based on scientific theory or prior experience, we are willing to assume parametric distributions for some of the transitions, but for others we prefer to let the data speak for themselves. In some cases we may even have *no* data for some transitions, but are willing to assume particular parametric distributions based on expert opinion (perhaps with a range of parameter values, yielding a range of estimates for the first passage distribution).

For example, in open queueing systems with random arrivals, the interarrival time distribution is often very close to exponential(λ), with λ the mean arrival rate, and it is standard to use this assumption (Kleinrock 1975); whereas the service time distribution is typically unknown and not well approximated by an exponential distribution. Thus it makes sense to use the exponential arrival assumption, while dealing with the service time distribution nonparametrically.

Little more needs to be said, since the computational details are perfectly straightforward. By viewing a parametric distribution function as the limit of EDFs from arbitrarily large samples, all the results associated with the first passage sample theorem (Section 2.1.2) continue to hold. Likewise, since an exact transform is trivially a strongly consistent estimator of itself, the convergence results of Section 2.1.3 also hold.

The additional computational overhead for nonparametric flowgraphs stems from the complexity of empirical transforms; the first derivative may have tens of thou-

sands of exponential terms, and for any inversion method either the transform or its derivatives must be evaluated many times for each plot point. In contrast, exact transforms for most of the commonly encountered distributions are quite simple, thus the overhead is significantly reduced for each empirical transform replaced by an exact transform. Accuracy is improved, assuming the hypothesized distribution family is correct.

One important use of nonparametric methods is to act as a check on the validity of results from a parametric model. In the flowgraph context, a mismatch between the parametric and nonparametric results may be analyzed further by a stepwise process. Starting from the fully parametric model, as each transition transmittance is replaced by an empirical version the result is examined in order to isolate the transition or transitions where the assumed parametric model is in error.

Section 3.1.4 provides an example, where parametric distributions are assumed for two out of the four transitions in the repairable redundant system example that is analyzed in a fully nonparametric way in Section 3.1. It illustrates both the improved accuracy and reduced overhead that semiparametric analysis may provide.

Some possible Bayesian semiparametric approaches to flowgraphs were discussed very briefly in Section 2.2.1 (p. 107). We have not pursued these in the current work.

2.5 Prior work and related methods

As noted in Section 2.1, there is considerable literature on empirical transforms. Except for one empirical saddlepoint paper discussed below, this prior work deals only with transforms of a single distribution, and focuses primarily on convergence results and the asymptotic distribution of the error introduced by the sample average process, not on inversion of the transforms. In a few cases, e.g., (Parzen 1962,

p. 1070) and (Silverman 1986, Section 3.5), the empirical CF is used to facilitate kernel smoothing for single-sample density estimation. Parzen uses the ECF only for theoretical purposes, Silverman inverts it using the fast Fourier inverse in order to reduce the computation required by kernel smoothing.

The first use of the empirical saddlepoint approxmation was in (Davison & Hinkley 1988). They considered an iid sample $\{x_1, \ldots, x_n\}$ from a distribution F, which we will assume here has density f. As opposed to our use so far of the saddlepoint approximation for the the density of X_i, they used it as originally intended, to approximate the sampling density f_n of the mean $n^{-1}\sum_i x_i$. For this purpose the approximation formula is (2.30) (p. 95), in which they substituted the empirical CGF $\tilde{\mathcal{K}}$ for $\mathcal{K}$. Their motivation was to use $\tilde{f}_n$ to compute, e.g., a confidence interval for the mean, which they showed is a good approximation to the bootstrap confidence interval obtained by resampling $\{x_1, \ldots, x_n\}$. The empirical saddlepoint method is advantageous in this case because it is simpler to compute than the bootstrap estimate.

Davison and Hinkley pointed out but did not prove the fact that using the ECGF, no solution for the saddlepoint equation exists for $\hat{s} < x_{(1)}$ or $\hat{s} > x_{(n)}$ (cf. Theorem 2.2.2), though they asserted "this is of no practical import" for the applications they had in mind. In density estimation for reliability and survival analysis, of course, it may be of great practical import, since it is often the tail behavior of distributions that is of greatest interest.

Though they did not conceptualize it this way, estimating the distribution of the mean (or equivalently the sum $\sum_{i=1}^n X_i$) can be represented as a flowgraph problem, as shown in Figure 2.26 where $\mathcal{M}(s)$ is the common MGF of the X_i. The $1 \to n+1$ first passage distribution is the distribution of $\sum_{i=1}^n X_i$. The equivalent bootstrap procedure would be to iterate summing sample holding times from the individual transitions and develop estimates based on the samples. (In this case,

because the holding time distributions are identical it would suffice to resample the original sample, with replacement, n times for each first passage sample.)

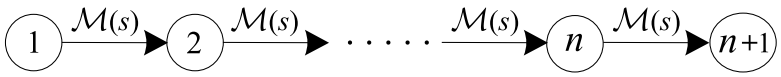


Figure 2.26: Flowgraph representing the sum of n iid random variables

Butler and Bronson (2002) adopted a similar approach to complex flowgraphs with feedback loops, using empirical MGFs and the saddlepoint approximation to achieve a result equivalent to bootstrapping from a sample of first passage times in a general semi-Markov process. In contrast to Davison and Hinkley, they used the saddlepoint density approximation (2.26) since the various transition distributions were not identical. They provided no proofs of convergence results like those of Sections 2.1.2 and 2.1.3 here, which must be assumed for the validity of the approximations they used. They also did not address the question of the range over which the empirical saddlepoint equation can be solved (Theorem 2.2.2), perhaps because they were primarily interested in specific percentiles of the distribution which may not have required going outside this range. Most importantly, their examples were restricted to distributions for which the saddlepoint density approximation is reasonable accurate; in general, as shown by Theorem 2.2.1, one cannot assume this accuracy for either parametric or nonparametric methods.

With the exception of (Butler & Bronson 2002), prior work in flowgraphs and semi-Markov processes generally has used parametric models for transforms. This dissertation goes beyond the previous work in the respects mentioned above, in handling censored data, in using the Fourier series inversion, and in recognizing the need for a systematic investigation of errors in the approximations.

For a system with transient operational states and a set of absorbing failure states, Limnios and Oprişan (2001, Proposition 5.3) derived a renewal equation giving the

reliability function $R_i(\tau)$ starting in a given state i. If we define j as the single failure state, then $1 - R_i(\tau)$ is the first passage distribution function $G_{ij}(\tau)$. Ouhbi and Limnios (1999) showed (Theorem 7) that this equation gives a consistent estimator of reliability when Pyke's nonparametric estimator $\hat{\mathbf{Q}}$ (Equation 1.1, p. 8) is substituted for $\mathbf{Q}$. They also proved (Theorem 8) a large-sample distributional result which depends on the unknown $\mathbf{Q}$, which one would guess might hold asymptotically after substituting $\hat{\mathbf{Q}}$. Their results may be used with an estimator for $\mathbf{Q}$ developed by Gill (1980a) for randomly right-censored data. However, the computational tractability of these methods is questionable. Solving Markov renewal equations is nontrivial, and may be very difficult for large numbers of states even when an exact parametric $\mathbf{Q}$ is known (Elkins & Wortman 2002; Kulkarni 1996, Chapt. 9). Typically transform methods are used, requiring numerical inversion, which is confronted by all the issues discussed so far. Even in the parametric case, we feel that flowgraphs provide a more manageable technique which is equivalent to solving the renewal equations (see Section 1.4.4, p. 27), and is both conceptually clear and computationally efficient. In the nonparametric case, we have found no prior work on solving the renewal equations that arise from semi-Markov processes.

Chapter 3

Example applications

This chapter contains examples that illustrate and evaluate the nonparametric flowgraph methodology. The scenarios used are realistic; data are generated by simulation to permit comparison with the true distributions. Each scenario will be described in enough detail to relate it to real-world problems where these methods could be applied.

In each case, results using different inversion methods are compared with each other and with results based on the true parametric distributions.

3.1 Repairable redundant system

In this section we revisit the reliability analysis example that was briefly discussed in Section 1.1. The first version of the analysis uses fairly large samples of transition data, consistent with study periods of a year or more in a well controlled field environment. The example is worked with both uncensored and censored data. Section 3.1.3 contains a version of the analysis based on small uncensored samples (with

sizes about 10% of the original sample sizes). Section 3.1.4 presents a semiparametric analysis (see Section 2.4) where we assume parametric exponential distributions for the $1 \rightarrow 3$ and $2 \rightarrow 3$ transitions and estimate the parameter using maximum likelihood, while using only the data for the $1 \rightarrow 2$ and $2 \rightarrow 1$ transitions.

Figure 3.1 shows the states of the system under study. Edges $i \rightarrow j$ are labeled with their transmittances $p_{ij}\mathcal{T}_{ij}(s)$ where $\mathcal{T}$ can be the LT or MGF. The system is composed of two units operating in parallel, and is functional if at least one of the units has not failed. A failing unit can be repaired; system failure occurs if both fail simultaneously, or if the second unit fails while the first is being repaired. Interest lies in estimating the probability distribution, or moments of the distribution, for the time of first passage from state 1 to state 3.

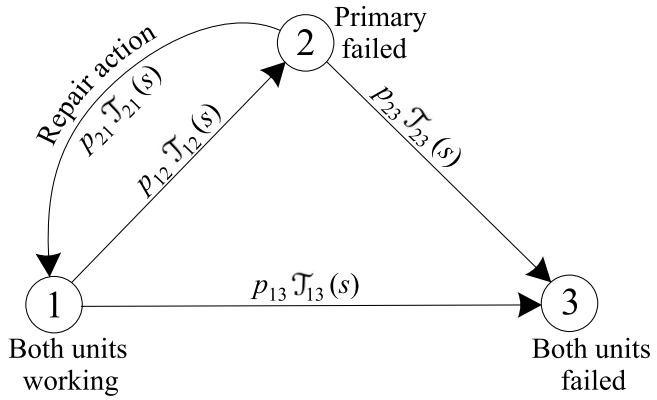


Figure 3.1: Three-state repairable redundant system model

It is assumed that each machine is independently capable of supporting whatever "load" is required for full system operation, and that the capability of each machine is independent of the number of failures that have occurred (perfect repair). Situations where the load is redistributed after a failure (Kvam & Peña 2005) can, in principle, be modeled with flowgraphs by adding a state to represent each post-failure redistribution; likewise we can handle imperfect repair and other departures

from time-homogeneity (Peña 2006) by proliferating states. However, for empirical flowgraphs the computational burden becomes excessive for "dynamic" problems of significant size, and one seldom has enough data in such cases to support a completely empirical analysis. In practice, systems often do (at least to a first approximation) satisfy independence and time-homogeneity with a reasonable number of states, and the method described here is applicable.

The $1 \to 2$ and $2 \to 1$ links each aggregate two transitions: primary fail/repair, and backup fail/repair; state 2 summarizes two states, primary failed and backup failed. The primary and backup units are considered identical, so the $1 \to 2$ transition occurs if either one fails, and the $2 \to 1$ transition occurs when the failed unit is repaired.

Both units are assumed to have independent exponential(λ_1) failure distributions, so with failure times T_1 and T_2 for the first and second unit, the distribution for T, the time for the $1 \to 2$ transition, is given by

$$\begin{aligned} P(T \le t) &= 1 - P(T > t) = 1 - P(T_1 > t, T_2 > t) \\ &= 1 - \exp(-\lambda_1 t)\exp(-\lambda_1 t) = 1 - \exp(-2\lambda_1 t). \end{aligned}$$

Thus the distribution of T is exponential($2\lambda_1$). The $1 \to 2$ transition includes the case of simultaneous random failures in both units, which should take the system to state 3; however, the probability of this is negligible and we ignore it. The direct $1 \to 3$ transition is presumed to result from common-cause failures, e.g., power outages, not from simultaneous random failures. The distribution for the $1 \to 3$ transition is modeled as exponential(λ_2). The repair time distribution is modeled as gamma(α, β), independently of which machine has failed. Specifically, taking the time unit as minutes, $\lambda_1 = 0.0002778$, corresponding to a mean time to failure (MTTF) of *either* unit of 30 hours (1.25 days); $\lambda_2 = 0.00002315$, MTTF of 30 days; $\alpha = 2$, $\beta = 180$ for a mean time to repair (MTTR) of 6 hours. Transition probabilities

are $p_{12} = .96$, $p_{13} = .04$, $p_{21} = .907$, and $p_{23} = .093$ (these were computed from the true distributions using competing risk ananysis, and for the sake of simplicity are assumed to be accurately estimated from the observed data).

The Mason's rule solution for the flowgraph in Figure 3.1 was derived in Section 1.3.1 (Equation 1.2, p. 16) as

$$\mathcal{T}(s) = \frac{p_{01}\mathcal{T}_{01}(s)p_{12}\mathcal{T}_{12}(s) + p_{02}\mathcal{T}_{02}(s)}{1 - p_{01}\mathcal{T}_{01}(s)p_{10}\mathcal{T}_{10}(s)}. \tag{3.1}$$

$\mathcal{T}$ in this case will be the Laplace transform for the EULER inversion and the MGF for the saddlepoint inversion. For comparison with nonparametric estimates, the true density for the $1 \to 3$ first passage distribution was computed in Mathematica by analytic inversion of the exact Laplace transform.

3.1.1 Results for uncensored Data

The model was analyzed using uncensored samples with 270, 25, 247, and 23 observations sampled from the $1 \to 2$, $1 \to 3$, $2 \to 1$, and $2 \to 3$ transition time distributions, respectively. Plots of the sample histograms and true holding time pdfs are shown in Figure 3.2.

EULER approximations for the density and CDF of the $1 \to 3$ first passage distribution are shown in Figure 3.3. Error statistics for these and the empirical saddlepoint approximation are displayed in Tables 3.1 and 3.2. The exponentially smoothed EULER result is better than the result from the modified EULER algorithm, but takes quite a bit longer to compute; on a 700 MHz processor, generating the plot points took 304 seconds for EULER, versus 89 seconds for modified EULER.

The first passage density is quite close in shape to a gamma density, so by Daniels' Conjecture (p. 96) we would expect a good result for the saddlepoint method, and

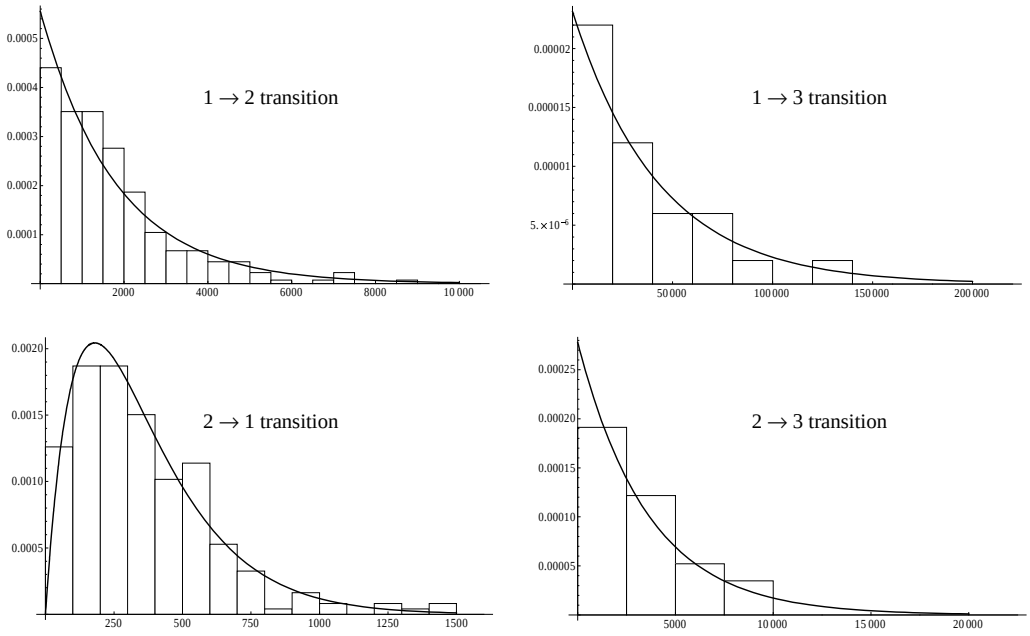


Figure 3.2: True pdfs and sample histograms for transitions in the repairable system example.

it does not disappoint. Figure 3.4 shows the empirical saddlepoint approximations to the pdf and CDF; Tables 3.1 and 3.2 display the error statistics.

The empirical saddlepoint is much slower than the Fourier series inversion: generating 1,000 plot points takes about 12 minutes. An adaptive algorithm would improve this significantly; because the saddlepoint result is fairly smooth, in principle it is possible to approximate the curvature (using the second derivative) and plot fewer points where the density is nearly linear. Looking at this and other potential performance improvements is a future research objective.

Table 3.1 shows the global error statistics, and Table 3.2 compares percentiles of the distribution for the approximations. Only smoothed EULER, not modified EULER, is shown in Table 3.2. The parametric saddlepoint result is shown for

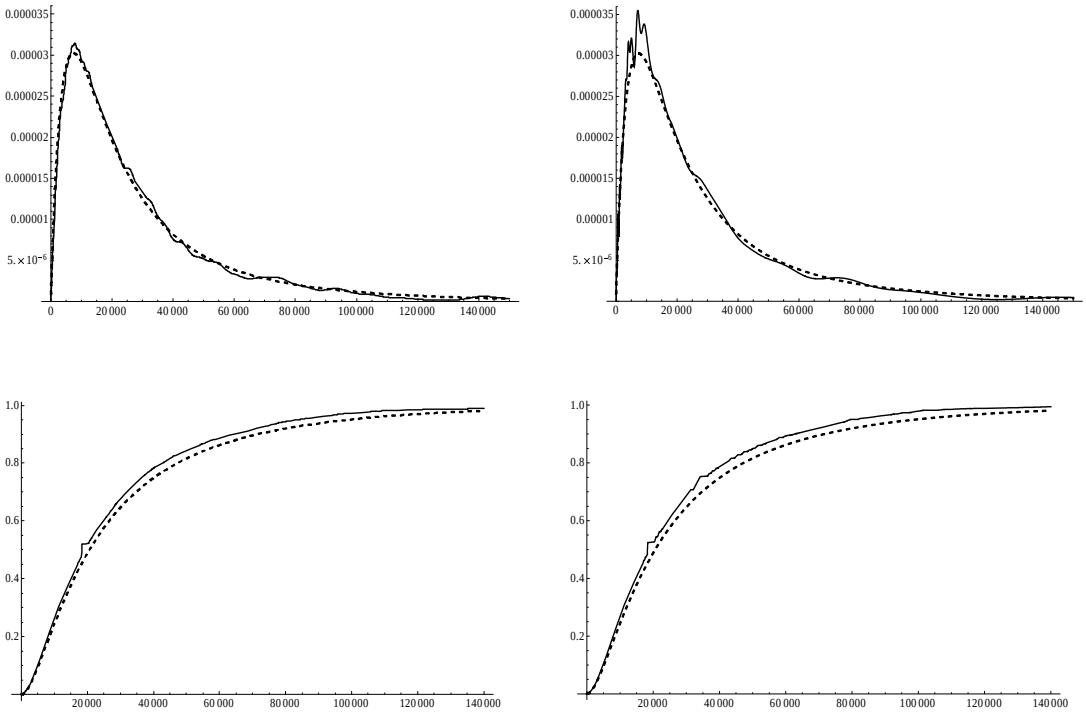


Figure 3.3: Approximations for the $1 \to 3$ first passage, uncensored data. Top row: pdf, smoothed EULER (left) and modified EULER; bottom row, EULER (left) and modified EULER CDF approximations; true curves are dashed.

comparison to the empirical saddlepoint. "Rel. error" is the error in the percentile relative to the true value. The poor results in the right tail of the distribution for EULER and the empirical saddlepoint are most likely a consequence of sampling,

Method	IAE (pdf)	K-S (CDF)
Smoothed EULER	0.06542	0.06456
Modified EULER	0.08216	0.06829
Empirical saddlepoint	0.07273	0.11580

Table 3.1: $1 \to 3$ first passage distribution, uncensored data: integrated absolute errors (IAE) for pdf and Kolmogorov-Smirnov (K-S) statistics for CDF.

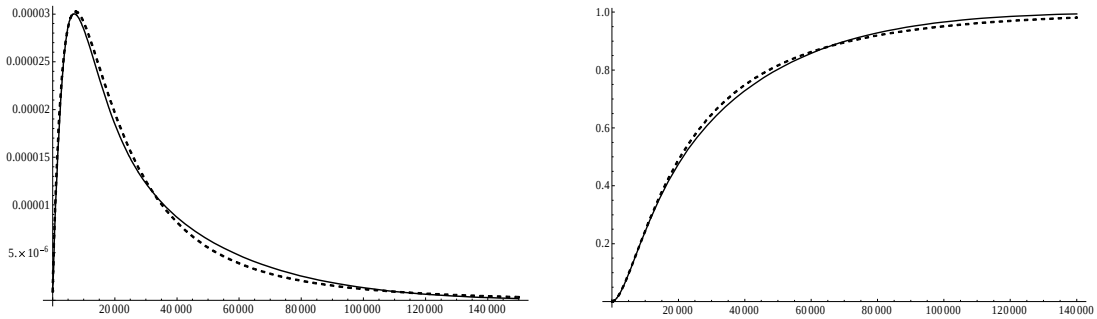


Figure 3.4: Empirical saddlepoint approximations for the $1 \rightarrow 3$ first passage, uncensored data. Left: pdf, right: CDF; true curves are dashed.

Percentile	Exact	Smoothed EULER	Rel. error	Parametric saddlepoint	Rel. error	Empirical saddlepoint	Rel. error
5	3,295	3,313	.006	3,312	.005	3,423	.039
25	10,177	9,643	.052	10,366	.019	10,352	.022
50	20,554	19,345	.059	21,487	.046	21,478	.045
75	40,176	36,356	.095	43,805	.090	42,715	.063
95	98,975	83,325	.158	105,859	.070	90,057	.090
99	167,090	138,867	.169	172,461	.032	126,565	.225

Table 3.2: $1 \rightarrow 3$ first passage distribution percentile results for uncensored data.

which is unlikely to produce points in the extreme tail. This loss of information in the tails is a general problem with nonparametric methods based on samples.

Using the procedure described in Section 2.3, an approximate 95% bootstrap confidence interval for the mean (MTTF) is (24,140, 34,699), and for the standard deviation, (20,387, 35,999) (values are in minutes). Exact values, computed from the analytic transform inversion, are $\mu = 31{,}650$, $\sigma = 34{,}052$.

The *hazard rate* $h(\tau)$ (sometimes called the failure rate), much used in reliability and survival analysis, is the instantaneous failure rate at time τ, conditional on

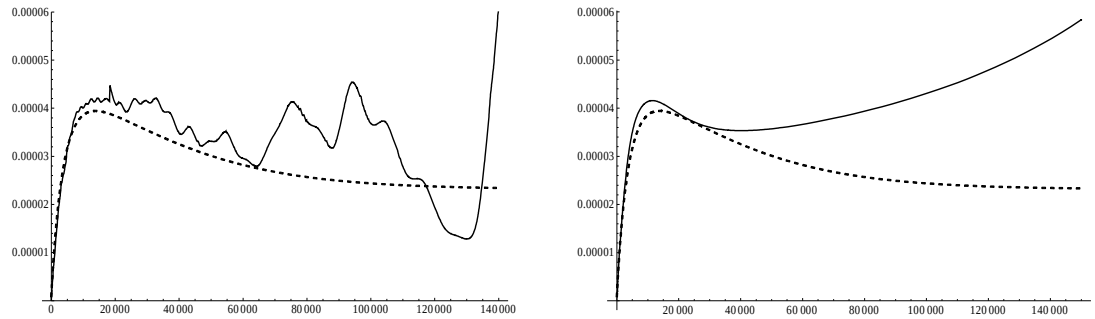


Figure 3.5: Hazard rate approximations for the $1 \to 3$ first passage, uncensored data. Left: EULER, right: empirical saddlepoint; true hazard rate is dashed.

survival to that point. Where $f(\tau)$ is the pdf of the failure distribution, $F(\tau)$ is the CDF, and $S(\tau) = 1 - F(\tau)$ is the survival function,

$$h(\tau) = \lim_{\Delta\tau \to 0} \frac{S(\tau) - S(\tau + \Delta\tau)}{\Delta\tau S(\tau)} = \frac{f(\tau)}{S(\tau)}. \tag{3.2}$$

The hazard rate is particularly sensitive to errors in density approximation, since the approximation is used both in the numerator and denominator of (3.2). This is clearly shown in Figure 3.5. Both approximations represent $h(\tau)$ with reasonable accuracy only up to about 30,000 minutes (close to the mean). The EULER approximation, though noisy, is qualitatively more accurate in that it correctly indicates that the hazard rate is decreasing after 30,000 minutes, whereas the saddlepoint approximation incorrectly shows it increasing. Past about 60,000 minutes the EULER approximation ceases to give any meaningful information.

3.1.2 Results for censored Data

This scenario is identical to that of Section 3.1.1 except for censoring. The assumption here is that all transitions leading to "both units failed" (state 2) are com-

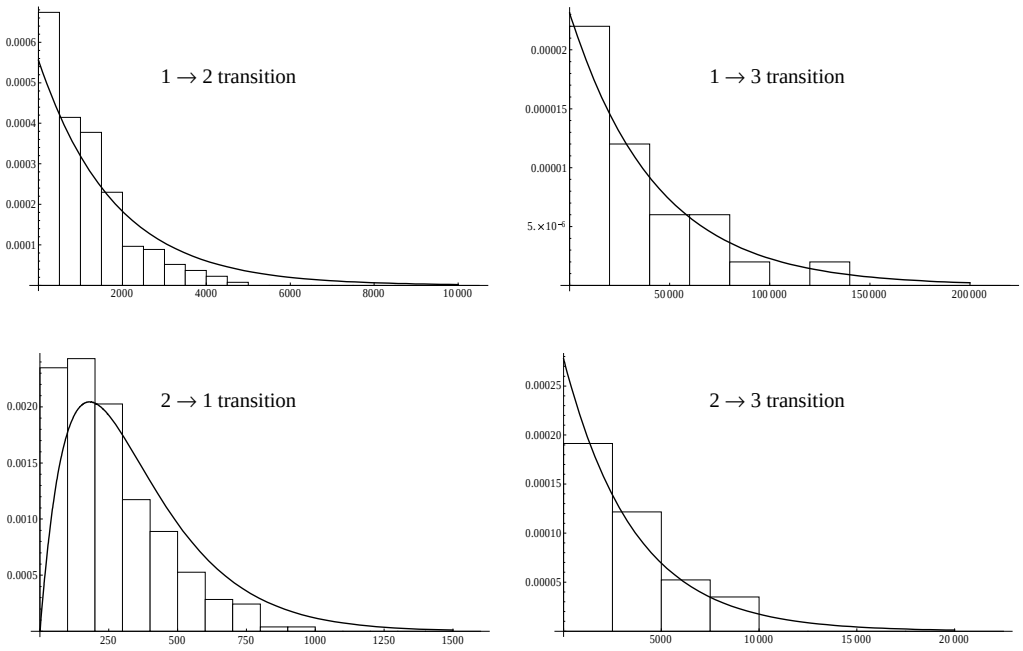


Figure 3.6: True pdfs and sample histograms for censored samples in the repairable system example.

pletely observed, but observations for the $1 \to 2$ and $2 \to 1$ transitions may be right censored (in the sample for this case 35% were censored for each transition). Independent exponential(λ_C) distributions were used to determine the censoring times. (This process is described in Appendix A.2.1, p. 170.) Histograms of the samples are plotted against the true holding time distributions in Figure 3.6; censoring of the longer times for the $1 \to 2$ and $2 \to 1$ transitions is evident.

Adjusted transforms were computed from the censored data as described in Section 2.1.4 (p. 2.1.4), and the remainder of the empirical flowgraph method was unchanged.

Visually, the density and distribution function approximations are very similar to the uncensored case and are not shown. Table 3.3 shows error statistics, Table

Method	IAE (pdf)	K-S (CDF)
Smoothed EULER	0.09523	0.03226
Modified EULER	0.11698	0.15515
Empirical saddlepoint	0.15945	0.13716

Table 3.3: $1 \rightarrow 3$ first passage distribution, censored data: integrated absolute errors (IAE) for pdf and Kolmogorov-Smirnov (K-S) statistics for CDF.

Percentile	Exact	Smoothed EULER	Rel. error	Empirical saddlepoint	Rel. error
5	3,295	3,284	.003	2,646	.197
25	10,177	9,763	.041	8,045	.201
50	20,554	18,863	.083	16,743	.185
75	40,176	34,880	.132	36,438	.093
95	98,975	82,490	.167	83,827	.153
99	167,090	141,993	.150	124,418	.255

Table 3.4: $1 \rightarrow 3$ first passage distribution percentile results for censored data.

3.4 compares percentiles of the distribution for smoothed EULER and the empirical saddlepoint approximations with the exact percentiles. EULER results are close to the uncensored case, the saddlepoint result is somewhat worse.

3.1.3 Small sample results

The model was reanalyzed using uncensored samples with 30, 10, 30, and 10 observations sampled from the $1 \rightarrow 2$, $1 \rightarrow 3$, $2 \rightarrow 1$, and $2 \rightarrow 3$ transition time distributions, respectively. Plots of the sample histograms and true holding time pdfs are shown in Figure 3.7. It turns out that the information obtainable from the nonparametric analysis is not very accurate, but consider the parametric alternative: is it obvious from these samples that the $1 \rightarrow 2$ transition has an exponential holding time distribution? Or that the $2 \rightarrow 1$ transition distribution is gamma? Or even

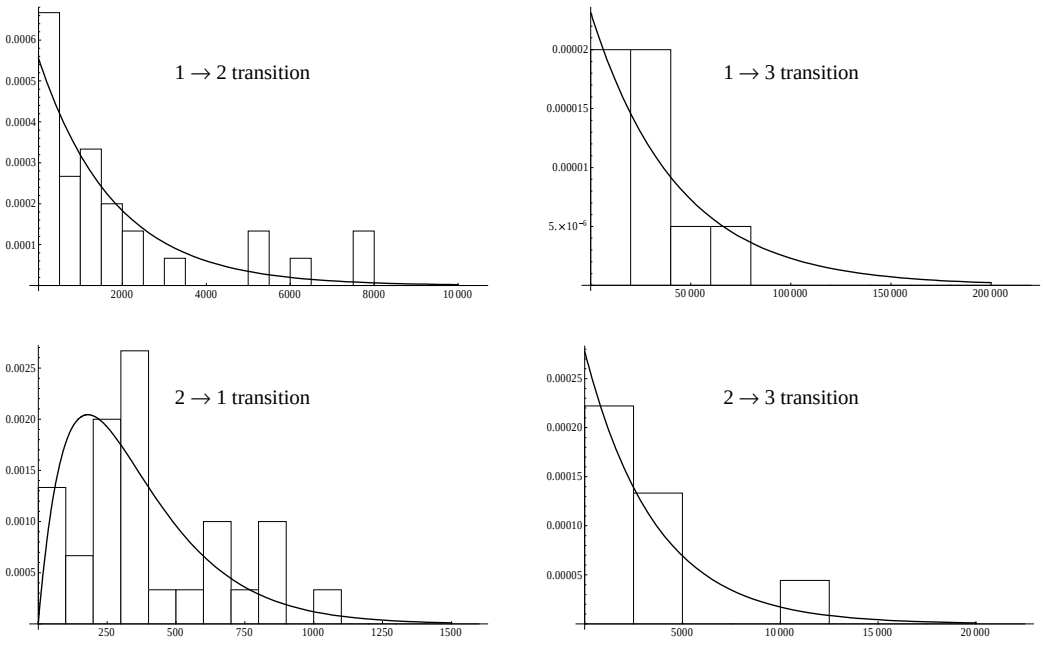


Figure 3.7: True pdfs and sample histograms for small samples in the repairable system example.

that they are both unimodal?

Processing time is reduced proportional to the smaller samples; generating the EULER plot points took 57 seconds with the small samples, versus 304 seconds with the large samples. Because the main advantage of the modified EULER algorithm is its faster processing, it was not used in this case.

Plots of the pdf and CDF approximations are shown for EULER in Figure 3.8, and for the empirical saddlepoint in Figure 3.9. Table 3.5 shows error statistics, and Table 3.6 compares percentiles of the distribution for smoothed EULER and the empirical saddlepoint approximations with the exact percentiles.

Although these results are not very accurate, they do provide a good general idea of the shape and scale of the distribution, which would certainly be helpful in

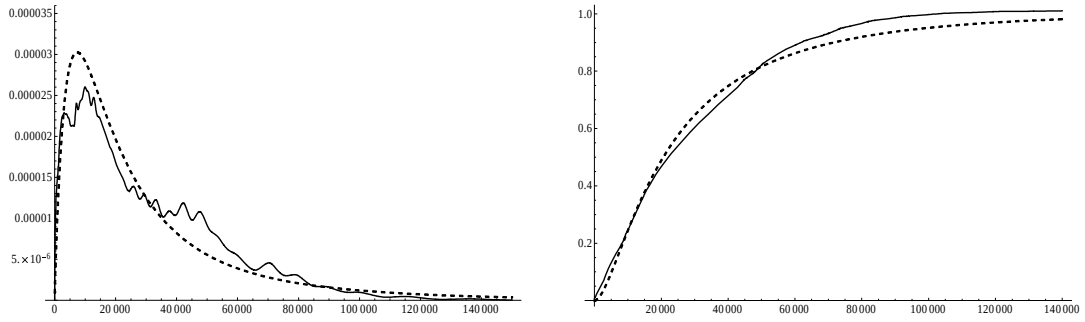


Figure 3.8: $1 \rightarrow 3$ first passage, small samples: smoothed EULER pdf approximation (left) and CDF approximation (right); true curves are dashed.

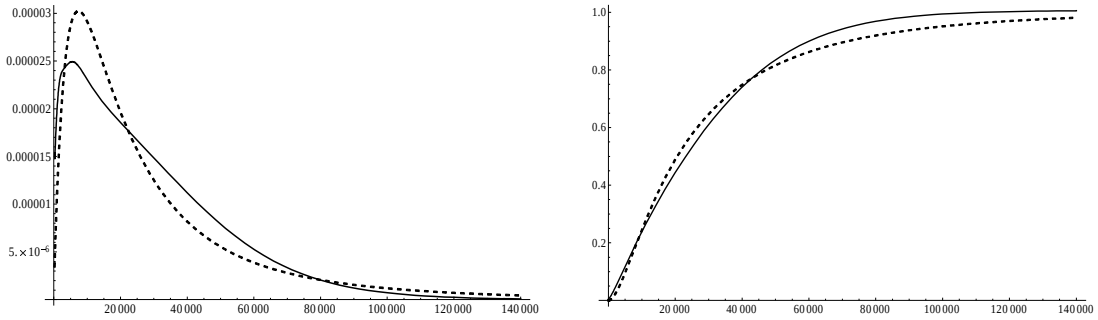


Figure 3.9: $1 \rightarrow 3$ first passage, small samples: empirical saddlepoint pdf approximation (left)and CDF approximation (right); true curves are dashed.

validating any alternative analysis, such as a parametric flowgraph model.

Method	IAE (pdf)	K-S (CDF)
Smoothed EULER	0.22866	0.05060
Empirical saddlepoint	0.21030	0.13195

Table 3.5: Repairable system example, small samples: integrated absolute errors (IAE) for pdf and Kolmogorov-Smirnov (K-S) statistics for CDF.

Percentile	Exact	Smoothed EULER	Rel. error	Empirical saddlepoint	Rel. error
5	3,295	1,962	.406	2,281	.301
25	10,177	10,378	.020	10,578	.039
50	20,554	22,448	.092	23,372	.137
75	40,176	44,006	.095	41,365	.030
95	98,975	78,254	.209	74,751	.245
99	167,090	104,773	.373	104,239	.376

Table 3.6: $1 \rightarrow 3$ first passage distribution percentile results for small samples.

3.1.4 Semiparametric analysis

Here we reanalyze the small sample case of Section 3.1.3 using the semiparametric technique described in Section 2.4 (p. 115). We model the $1 \rightarrow 2$ and $2 \rightarrow 1$ transitions nonparametrically, but assume parametric exponential(λ_{ij}) models for the $1 \rightarrow 3$ and $2 \rightarrow 3$ transitions. The λ_{ij} are estimated using the MLE $1/\bar{t}_{ij}$, where $\bar{t}_{ij}$ is the sample mean. This yields $\lambda_{13} = 3.36695 \times 10^{-5}$ (versus the true value 2.31481×10^{-5}), and $\lambda_{23} = 1.58113 \times 10^{-4}$ (versus the true value 2.77778×10^{-4}).

Processing time is reduced further here, because evaluating the parametric transforms is much faster than evaluating the corresponding empirical transforms. Generating the EULER plot points took 18 seconds, versus 57 seconds with the empirical small samples and 304 seconds with the large samples.

Plots of the pdf and CDF approximations are shown for EULER in Figure 3.10, and for the empirical saddlepoint in Figure 3.11. Table 3.7 shows error statistics, and Table 3.8 compares percentiles of the distribution for smoothed EULER and the empirical saddlepoint approximations with the exact percentiles. As we would expect, these results are significantly better than the comparable results for the pure nonparametric analysis with small samples. They are not as good as the pure nonparametric analysis with large samples.

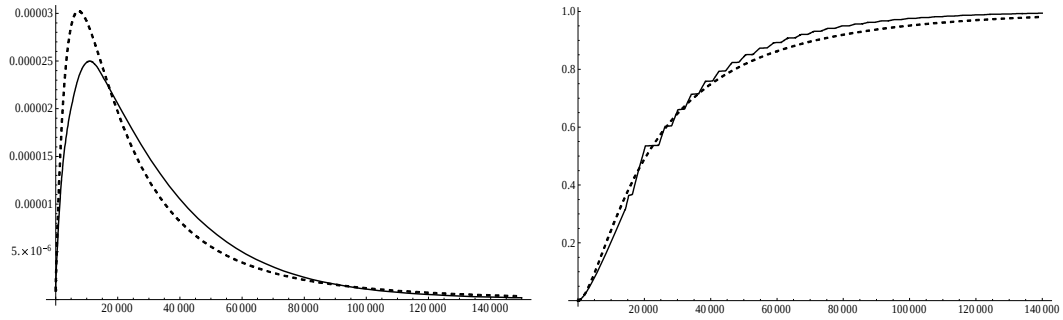


Figure 3.10: $1 \rightarrow 3$ first passage, semiparametric: smoothed EULER pdf approximation (left) and CDF approximation (right); true curves are dashed.

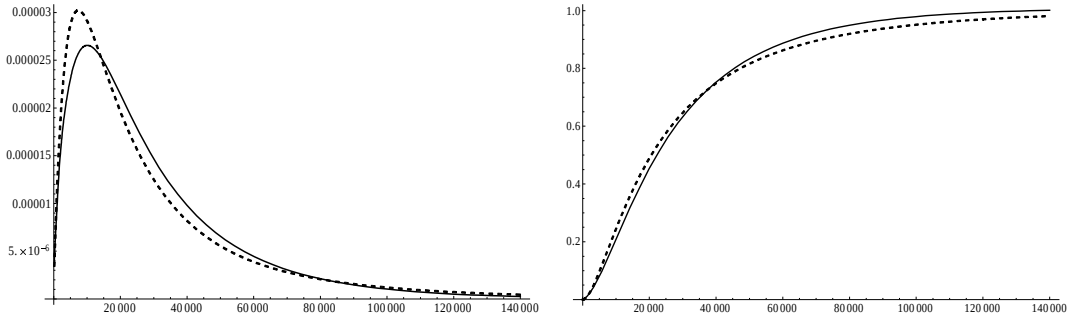


Figure 3.11: $1 \rightarrow 3$ first passage, semiparametric: empirical saddlepoint pdf approximation (left)and CDF approximation (right); true curves are dashed.

3.2 Mixture of wearout and random failures

This example is drawn from (Briand *et al.* 2008). A complex component is subject to two types of failures: random failure with exponential(λ) distribution (e.g., failure

Method	IAE (pdf)	K-S (CDF)
Smoothed EULER	0.18855	0.04595
Empirical saddlepoint	0.12816	0.05595

Table 3.7: Repairable system example, semiparametric analysis: integrated absolute errors (IAE) for pdf and Kolmogorov-Smirnov (K-S) statistics for CDF.

Percentile	Exact	Smoothed EULER	Rel. error	Empirical saddlepoint	Rel. error
5	3,295	3,680	.117	3,601	.093
25	10,177	11,708	.150	11,701	.150
50	20,554	20,337	.011	22,401	.090
75	40,176	37,385	.069	40,301	.003
95	98,975	81,703	.175	83,901	.152
99	167,090	126,403	.243	127,800	.235

Table 3.8: $1 \to 3$ first passage distribution percentile semiparametric results.

of an electronic subcomponent), and wearout of a mechanical part centered at a time μ after being placed in service. Briand *et al.* used a normal(μ, σ^2) distribution for wearout, which in principle could lead to negative failure times; to avoid this and for analytic tractability we use as a wearout failure distribution gamma(α, β) with $\alpha\beta = \mu$ and $\alpha\beta^2 = \sigma^2$. For this example the parameter values are $\lambda = 0.00333333$, $\alpha = 100$, and $\beta = 3$, thus $\mu = 300$ and $\sigma = 30$. The mixing probabilities are .4 for the gamma and .6 for the exponential. The flowgraph for this model is shown in Figure 3.12; the $2 \to 4$ and $3 \to 4$ transitions are added just to provide a common endpoint, so $p_{24} = p_{34} = 1$ and $\mathcal{T}_{24}(s) = \mathcal{T}_{34}(s) \equiv 1$. Then the Mason's rule expression for the $1 \to 4$ first passage transmittance is

$$\mathcal{T}(s) = p_{12}\mathcal{T}_{12} + p_{13}\mathcal{T}_{13}.$$

Analysis of a simple mixture such as this can be done without flowgraphs, using classical statistical methods (Lindsay 1995). However, it is instructive to do a flowgraph analysis because it exposes some interesting issues. A more complex variant of this example, which *does* benefit from flowgraphs, is presented in Subsection 3.2.1.

Figure 3.13, on the left, plots the two failure densities and the resultant mixture. As discussed in Section 2.2.3 (p. 95 ff.), this is the sort of density that is poorly approximated by the saddlepoint method. The right side of Figure 3.13 illustrates

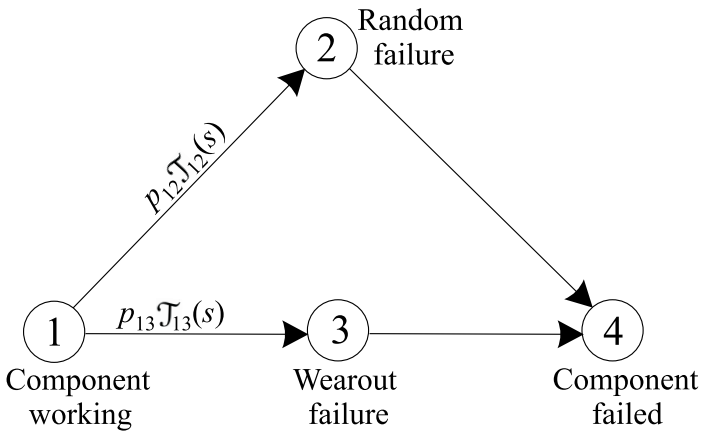


Figure 3.12: Flowgraph for wearout and random failure modes

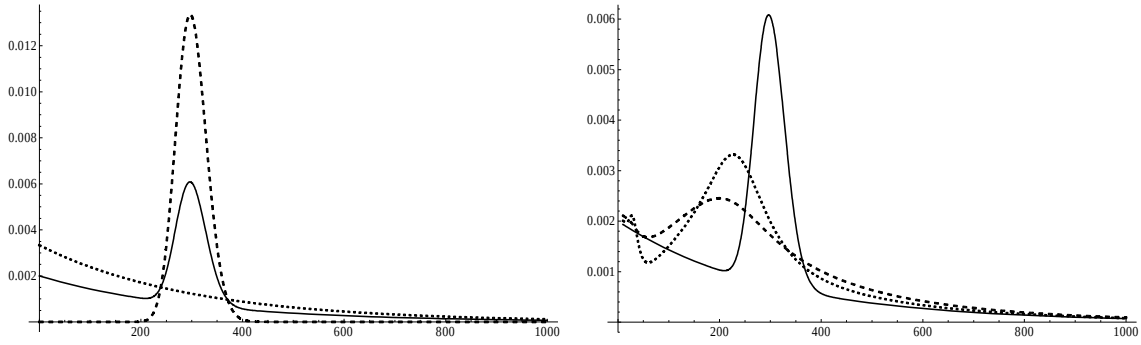


Figure 3.13: Left: Wearout failure density (dashed), random failure density (dotted), mixture density (solid line). Right: mixture density plotted with first-order (dashed) and second-order (dotted) parametric saddlepoint approximations.

this—even the second-order parametric saddlepoint (Equation 2.29, p. 94) is not a very good approximation. In contrast, the EULER inversion of the exact transform (not shown) is visually nearly perfect, with IAE = 0.0020.

For both the saddlepoint and Fourier series inversions, there is a second issue besides the accuracy of the saddlepoint for mixture densities. The sample sizes for

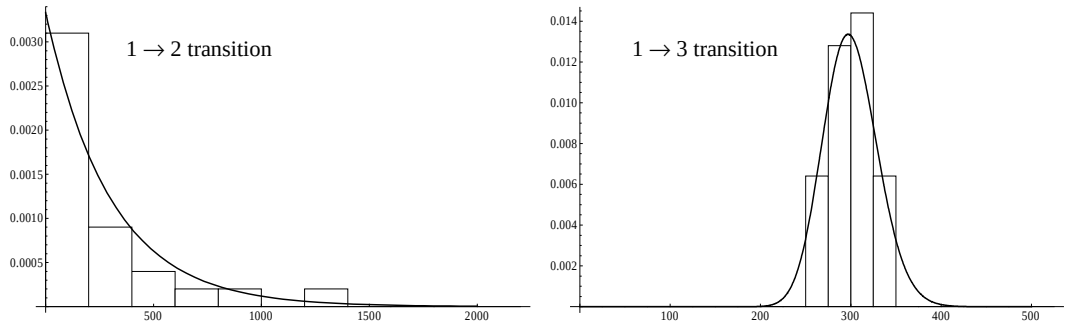


Figure 3.14: True pdfs and sample histograms for random (left) and wearout (right) failures.

this example are not large— 50 for the $1 \to 2$ transition, 25 for the $1 \to 3$ transition (samples are plotted in Figure 3.14). Recall from Section 2.2.2, p. 78, that in a flowgraph with loops and series of transitions the effective sample size is proportional to the product of the actual sample sizes. In this flowgraph, there are only parallel transitions, so the effective sample size is merely additive. This significantly reduces the potential accuracy of the inversions. In addition, the lack of feedback loops means that the saddlepoint equation cannot be solved for any value of s greater than the maximum failure time in either the $1 \to 2$ or $1 \to 3$ sample (see Theorem 2.2.2, p. 102), which is 1,264 in this case. This is near the limit of the effective support of the mixture density, so the pdf is set to zero past this point.

Figure 3.15 shows the pdf and CDF approximations computed by smoothed EULER and the empirical saddlepoint method, and Table 3.9 displays the error statistics. These results illustrate the value of looking at both approximations. The saddlepoint result correctly indicates that the mixture density has two modes, one at the origin, but in all other respects it is less accurate than EULER. EULER correctly places the larger mode of the density, and one can infer that there is a second mode at the origin, but the plot has too much noise to assess exactly how many

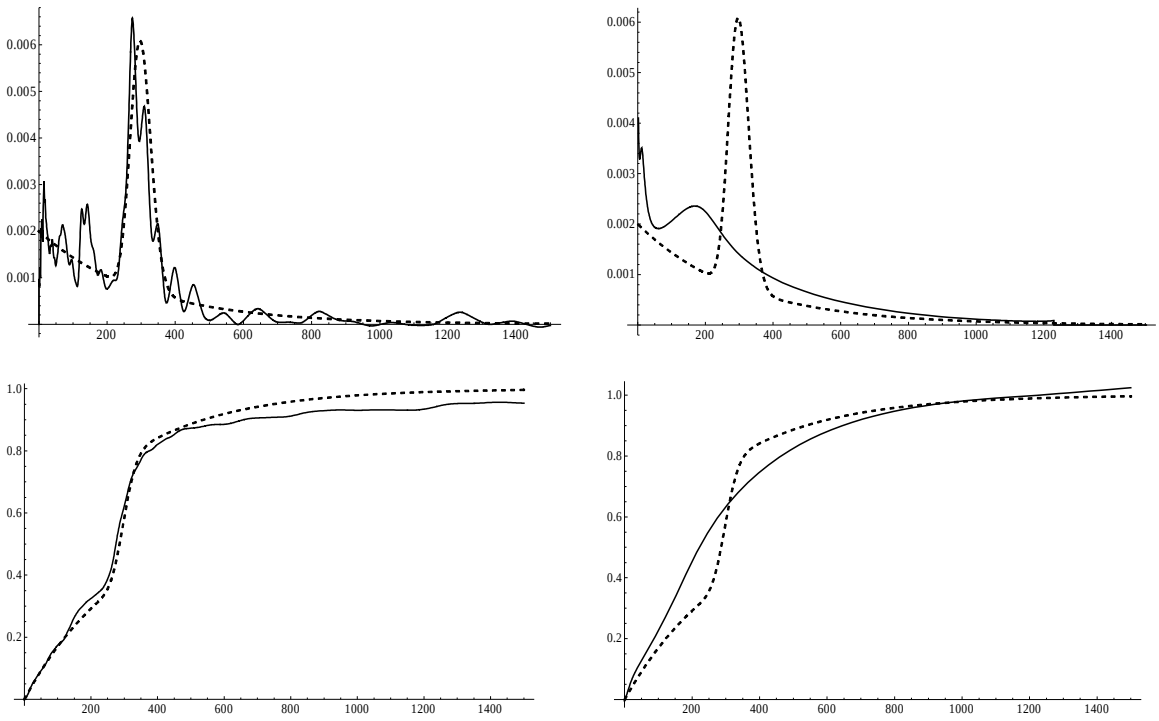


Figure 3.15: Top left: smoothed EULER approximation to wearout/random mixture density; top right: empirical saddlepoint approximation; bottom row shows corresponding CDF approximations; true curves are dashed.

modes there are. The EULER CDF approximation is clearly much more useful than the saddlepoint CDF. Comparing Figure 3.15 with Figure 3.13, it is obvious that the error in the empirical saddlepoint approximation is endemic to the saddlepoint density approximation, and not simply a consequence of sampling error.

3.2.1 A more complex mixture example

Figure 3.16 shows a variation on the wearout/random mixture example. In this case, we assume that a worn-out part can be instantly replaced with probability p_{31}, avoiding a failure; with probability $p_{34} = 1 - p_{31}$ there is no replacement available,

Method	IAE (pdf)	K-S (CDF)
Smoothed EULER	0.32772	0.06439
Empirical saddlepoint	0.62240	0.24546

Table 3.9: Wearout/random mixture example: integrated absolute errors (IAE) for pdf and Kolmogorov-Smirnov (K-S) statistics for CDF.

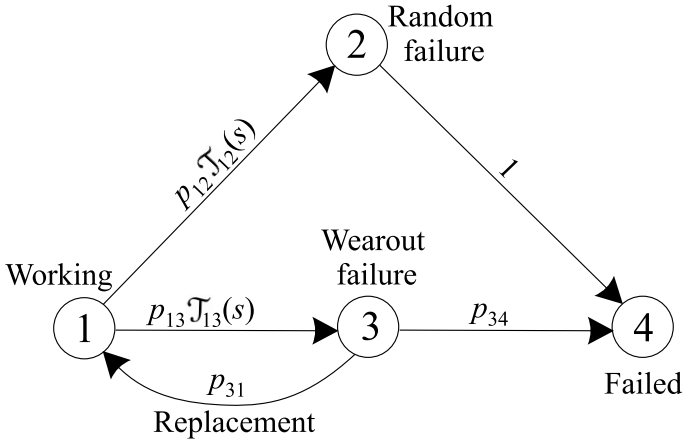


Figure 3.16: Flowgraph for wearout and random failure modes, with probabilistic immediate replacement after wearout.

resulting in failure. An example is the common classroom overhead projector: ideally, there is a spare bulb that can be swapped for a burned-out bulb almost instantly by moving a lever. However, it may be that the last user who experienced a burn-out did not get the bulb replaced, in which case the "spare" is also burned-out and the projector has failed. Because there is a feedback loop $1 \to 3 \to 1$, this is no longer a classic mixture problem, and we need a flowgraph model or some equivalent to solve it. The Mason's rule expression for the $1 \to 4$ first passage transmittance is

$$\mathcal{T}(s) = \frac{p_{12}\mathcal{T}_{12}(s) + p_{13}\mathcal{T}_{13}(s)p_{34}}{1 - p_{13}\mathcal{T}_{13}(s)p_{31}}.$$

We assume here that $p_{31} = p_{34} = .5$; otherwise all the parameters and the samples

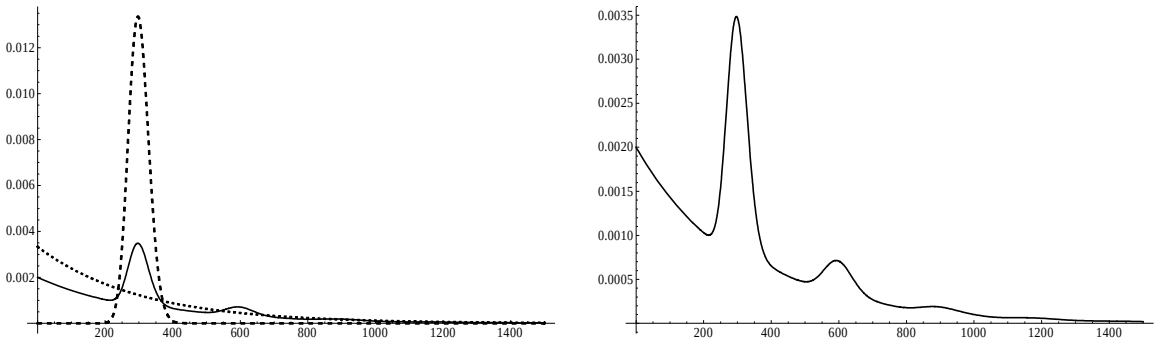


Figure 3.17: Left: wearout (dashed), random (dotted), and $1 \to 4$ first passage (solid) densities; right: enlarged view of $1 \to 4$ first passage density.

are the same as in the simple mixture case. What was the $1 \to 3$ sample for wearout failure now includes cases of successful replacement, as well as wearout resulting in system failure. Figure 3.17 shows the component failure densities and the $1 \to 4$ first passage density. At least four modes are visible in the $1 \to 4$ first passage density; there are actually an infinite number of modes, corresponding to failure after $0, 1, 2, \ldots$ successful replacements.

Figure 3.18 shows the pdf and CDF approximations computed by smoothed EULER and the empirical saddlepoint method; Table 3.9 displays the error statistics. The saddlepoint density result is worse than in the simple case, smoothing away all but the first two modes, and almost losing the second. Unlike in the simple case, because of the feedback loop the saddlepoint equation can be solved over the full support of the density on the right. The slight glitch on the left is because the minimum sample value is $1 > 0$. EULER does catch the third mode, but this is visible only because the true density is displayed; in practice, it's unlikely that it could be differentiated from the noise.

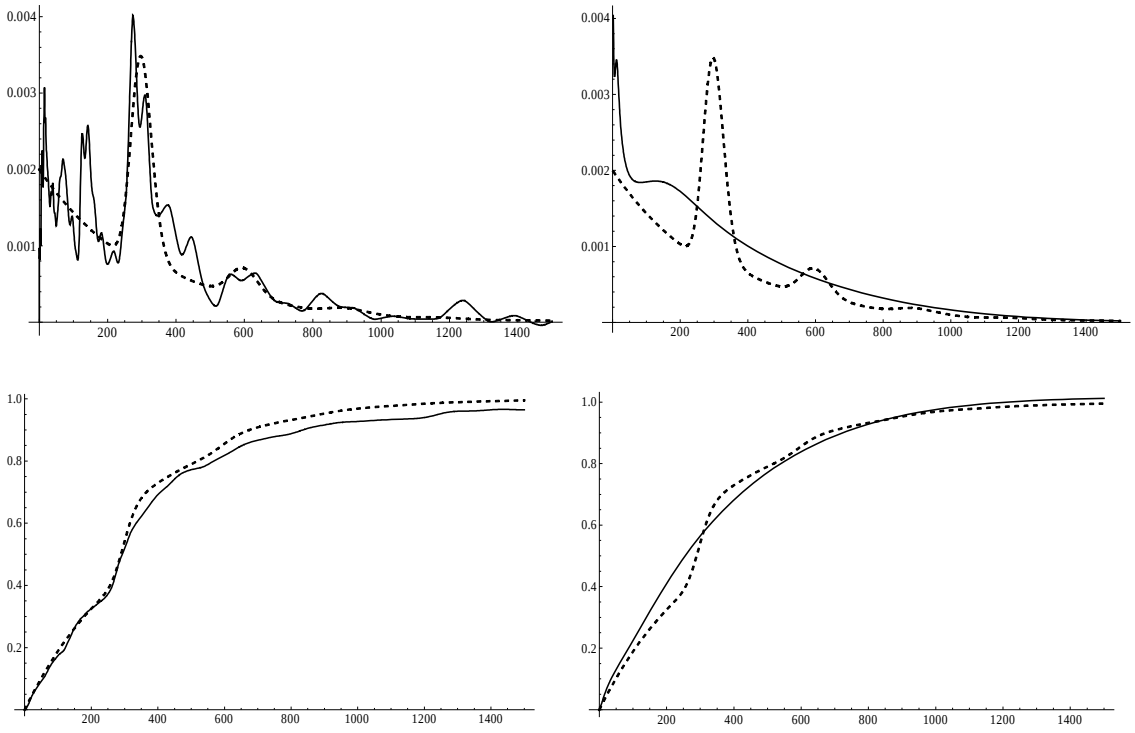


Figure 3.18: Top left: smoothed EULER approximation to wearout/random complex mixture density; top right: empirical saddlepoint approximation; bottom row shows corresponding CDF approximations; true curves are dashed.

Complex mixture with additional feedback loops

To amplify the point made in Section 2.2.2, that added complexity in a flowgraph may improve accuracy, we modified the complex mixture flowgraph (Figure 3.16) by

Method	IAE (pdf)	K-S (CDF)
Smoothed EULER	0.28566	0.04423
Empirical saddlepoint	0.34366	0.14603

Table 3.10: Wearout/random complex mixture example: integrated absolute errors (IAE) for pdf and Kolmogorov-Smirnov (K-S) statistics for CDF.

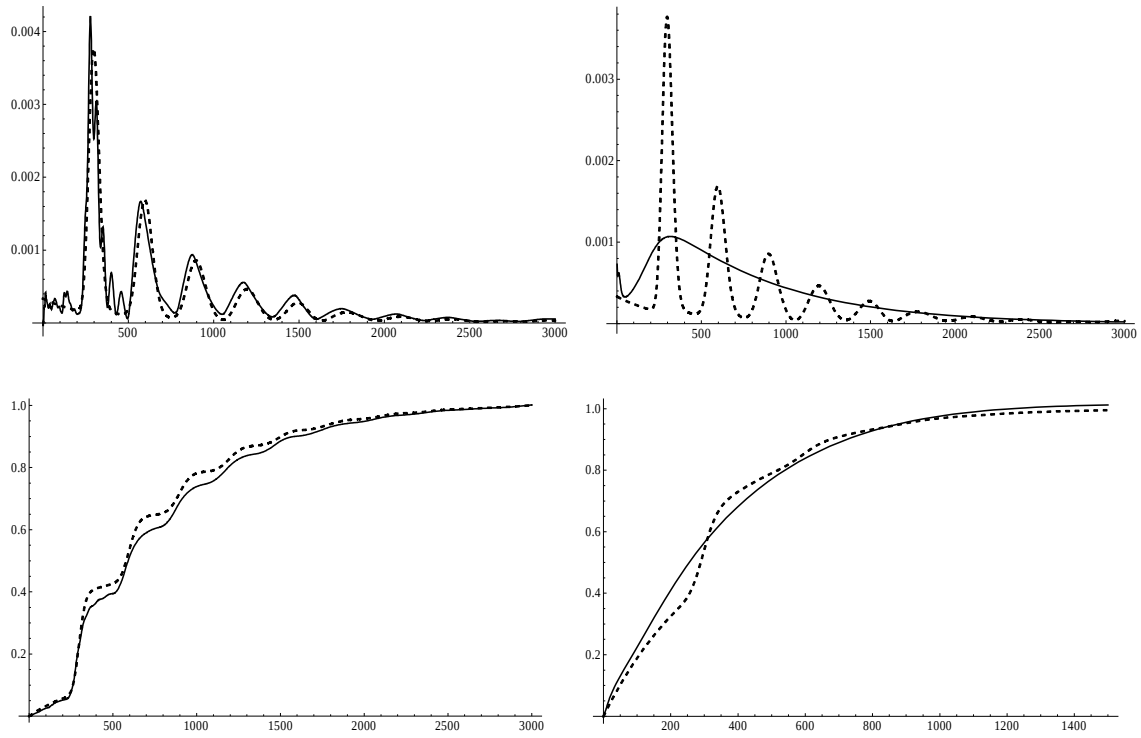


Figure 3.19: Smoothed EULER (left) and empirical saddlepoint (right) approximations for the complex mixture with additional loops; true curves are dashed. Compare Figure 3.18.

increasing p_{13}, the probability of wearout (as opposed to random failure) from .4 to .9, and increasing p_{31}, the probability of successful replacement, from .5 to .7. This causes the loop $1 \to 3 \to 1$ to be traversed many more times, and accentuates the later modes of the density.

Figure 3.19 shows the results. The EULER approximations are clearly improved, as expected. The saddlepoint results are not, but that is also expected, on account of the poor performance of saddlepoint methods generally for multimodal densities. For first passage densities of the type for which the saddlepoint density approximation tends to be accurate, accuracy would be improved by more loop traversals in the

flowgraph.

Summarizing, the flowgraphs discussed in this section are challenging because the samples are relatively small, they are not (except in the last example) effectively increased by loop traversal, and the density to be approximated is multimodal. (The problem of multimodal densities affects nonparametric estimation methods in general, not just flowgraphs.) Nevertheless, nonparametric flowgraph analysis still produces useful results.

3.3 Cumulative earthquake damage to structures

Cumulative damage is a process observed in many engineering applications; examples include fatigue cracking in metals and damage to structures as a result of repeated mechanical shocks. Markov process models are widely used for understanding and prediction of cumulative damage (Bogdanoff & Kozin 1985), as they have been found to offer a reasonable compromise between fidelity and mathematical tractability. The simplest such model is the finite or countable-state Markov chain, in which both states and transition times are discrete. This is not as limiting as it may seem, since even obviously continuous state variables such as the length of fatigue cracks can be usefully represented as sequences of discrete magnitudes (Bogdanoff & Kozin 1985; Pappas *et al.* 2001). In looking at structures and other complex mechanical systems, time is often naturally discretised by scheduled inspection intervals (Morcous 2006).

As a consequence of the Markov property, probability distributions for holding times between states in Markov process models are limited to be of "memoryless" type—geometric in discrete time, or exponential in continuous time. The greater generality of semi-Markov process models allows holding times to be arbitrarily distributed, but introduces conceptual and computational complexity that has limited their use in engineering applications.

The example presented here, a four-state model of cumulative damage to structures as a result of repeated seismic shocks, is from (Gusella 1998, 2000). We use it to illustrate the added power of semi-Markov models generally and flowgraphs in particular, and to show how the nonparametric flowgraph method can be used to detect a model specification error.

The flowgraph for the model is shown in Figure 3.20. States represent discrete damage categories, with state 4, collapse of the structure, being an absorbing state. From states 1-3 there is a probability of progressing to any state with a higher level

of damage, with transitions triggered by seismic events of a given intensity. The main engineering interest lies in being able to predict (in a probabilistic sense) the time elapsed from having an undamaged structure to collapse, or equivalently the probability of collapse within a given time interval. These results follow from an estimate of the $1 \to 4$ first passage distribution. Note that this flowgraph has self-transitions, which we normally avoid (see Section 1.2, p. 3); here we use them in order to be faithful to Gusella's presentation of the model.

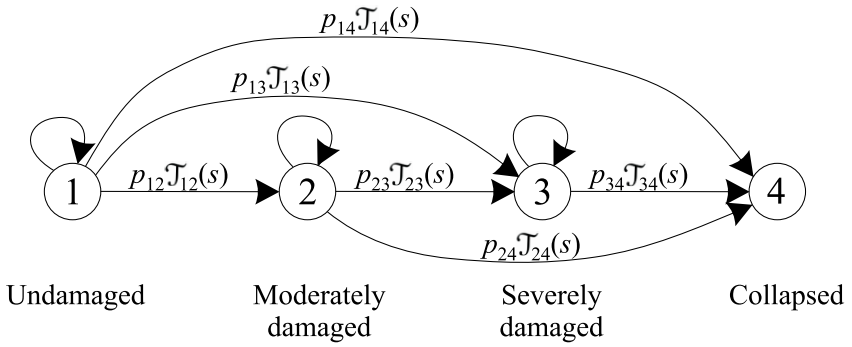


Figure 3.20: States and transitions for earthquake damage model

Gusella's model is hierarchical: seismic events (earthquakes of given intensity) are outcomes of a Poisson process with stationary parameter λ; given a seismic event, transition of a structure to a new damage state is given by a Markov chain with transition probability matrix P. In a Poisson process with mean λt events in time t the interarrival times are exponentially distributed with rate λ, so alternatively the model can be described as a semi-Markov process with the transition probability matrix of the embedded chain being P, and the holding time distributions being identically exponential(λ). Thinking of the process as semi-Markov facilitates generalization to non-identical, non-exponential holding times.

The process by which seismic events are propagated through the earth is enormously complicated (Richter 1958); thus any simple probabilistic model for earth-

quake occurrences and intensities will not be completely realistic. In particular, the adequacy of modeling series of seismic events as Poisson processes has been a subject of debate in the seismology literature (Vere-Jones 1970). Flowgraphs (and semi-Markov process models generally), as opposed to Markov process models such as Gusella's, offer the flexibility to accommodate arbitrary time distributions between seismic events.

The transition probability values in (Gusella 1998) are based on a physical model for the effect of ground acceleration on masonry structures, which is another very complicated process (Kanai 1983). We take the physical model at face value, since we are mainly interested in general methodological features of the model.

Let $\mathcal{T}_{ij}$ be the adjacent-state transition transforms, and $\mathcal{U}_{34}$ the transform of the $3 \rightarrow 4$ first passage distribution; then

$$\begin{aligned} \mathcal{U}_{34}(s) &= p_{33}\mathcal{T}_{33}(s)\mathcal{U}_{34}(s) + p_{34}\mathcal{T}_{34}(s) \\ &= \frac{p_{34}\mathcal{T}_{34}(s)}{1 - p_{33}\mathcal{T}_{33}(s)}. \end{aligned}$$

By continuing this recursion back to state 1 or by applying Mason's rule directly, and using the fact that every transition in this flowgraph has the identical transform $\mathcal{T}(s)$, the transform for the $1 \rightarrow 4$ first passage is given by

$$\begin{aligned} \mathcal{U}_{14}(s) = {} & \frac{p_{12}p_{23}p_{34}\mathcal{T}(s)^3}{(1 - p_{11}\mathcal{T}(s))\,(1 - p_{22}\mathcal{T}(s))\,(1 - p_{33}\mathcal{T}(s))} + \frac{p_{12}p_{24}\mathcal{T}(s)^2}{(1 - p_{11}\mathcal{T}(s))\,(1 - p_{22}\mathcal{T}(s))} + \\ & \frac{p_{13}p_{34}\mathcal{T}(s)^2}{(1 - p_{11}\mathcal{T}(s))\,(1 - p_{33}\mathcal{T}(s))} + \frac{p_{14}\mathcal{T}(s)}{1 - p_{11}\mathcal{T}(s)}. \end{aligned} \tag{3.3}$$

Gusella models four different categories of seismic events, characterized by their maximum ground acceleration. These are combined probabilistically to get an overall model for cumulative structural damage. For the purpose of illustration, only one

Method	IAE (pdf)	K-S (CDF)
Smoothed EULER	0.06272	0.03316
Empirical saddlepoint	0.13991	0.10767

Table 3.11: Cumulative damage example: integrated absolute errors (IAE) for pdf and Kolmogorov-Smirnov (K-S) statistics for CDF.

of the ground acceleration models is considered here (Gusella's A_{p4}) with parameter $\lambda = .0019$ for the Poisson arrival process and transition probability matrix

$$\begin{bmatrix} .1000 & .5333 & .2667 & .1000 \\ 0 & .3667 & .3000 & .3333 \\ 0 & 0 & .1000 & .9000 \\ 0 & 0 & 0 & 1.0000 \end{bmatrix}.$$

Sample data for the one holding time distribution were generated by simulation from an exponential(.0019) distribution. The sample size is $n = 100$, which is not unrealistic considering that thousands of seismogram records exist for most seismically active regions (Lee *et al.* 1988).

Approximation results for the $1 \to 4$ first passage distribution are shown in Figure 3.21, and the error statistics are in Table 3.11. As usual, the true pdf was determined by analytic inversion. Although not quite as smooth, the smoothed EULER result is clearly superior to the empirical saddlepoint in accuracy. This density, which is close in shape to a gamma, is the type that is approximated well by the saddlepoint method. The parametric saddlepoint approximation has an IAE of 0.011148 relative to the true distribution, and the parametric EULER result has IAE 2.3687×10^{-8}.

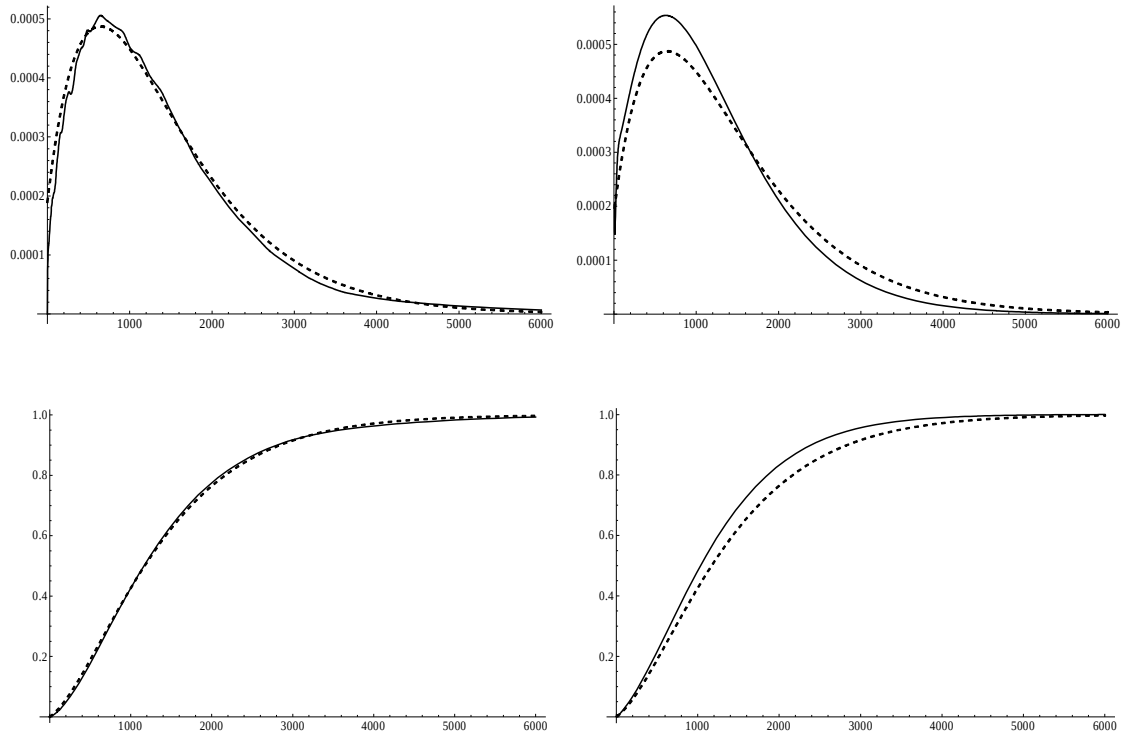


Figure 3.21: Top left: smoothed EULER approximation to $1 \to 4$ first passage density; top right: empirical saddlepoint approximation; bottom row shows corresponding CDF approximations; true curves are dashed.

3.3.1 Detecting a model error

Suppose that the assumption of a homogeneous Poisson arrival process (equivalent to an exponential distribution for arrival intervals) for seismic events is not valid. For example, suppose that the actual distribution of interarrival times is gamma(α, β) with $\alpha = 2.5$, $\beta = 211$. These values were chosen so that the mean interarrival interval is close to that of the hypothesized Poisson process in the Gusella model. (For a discussion of the plausibility of gamma-distributed arrivals, see (Takahashi *et al.* 2004), Section 3.2.) Figure 3.22 plots the gamma and exponential interarrival densities, and the densities for the $1 \to 4$ first passage distribution in the earthquake

damage model under the gamma and Poisson assumptions.

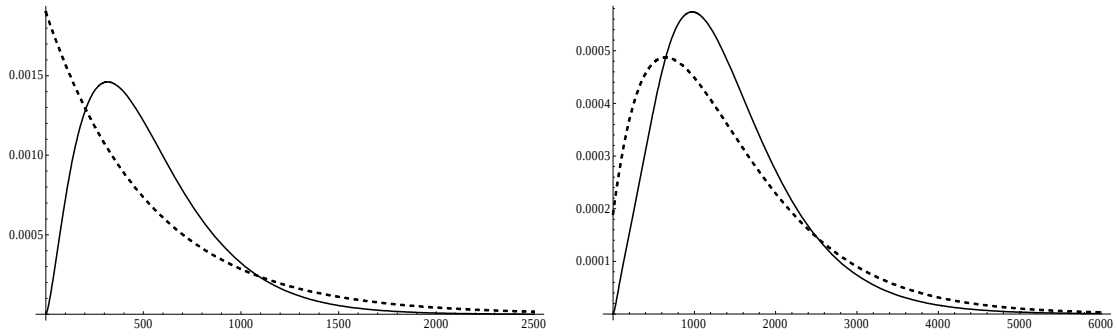


Figure 3.22: Left: Gamma pdf (solid) and exponential pdf (dashed) for seismic events. Right: $1 \rightarrow 4$ first passage density under the gamma model (solid) and exponential model (dashed).

To illustrate how the nonparametric method can detect the bad assumption in this parametric model, we use the following scenario. A sample of size 100, with sample mean $\bar{t} = 525.99$, was drawn from the gamma distribution. Applying the parametric flowgraph method with the assumption of exponential(λ) interarrival times, λ is estimated using maximum likelihood as $1/\bar{t} \approx .0019$ (the value in the Gusella model); the transform $\mathcal{T}(s)$ is computed accordingly, then Mason's rule is used to derive the $1 \rightarrow 4$ first passage transmittance (Equation 3.3, p. 145), which is inverted. For comparison, the $1 \rightarrow 4$ first passage density is estimated using the nonparametric method. The results shown below underscore the conclusion of (Gbur and Collins 1989), discussed on p. 37—the superiority of parametric methods is critically dependent on assuming the correct model.

Figure 3.23 plots nonparametric and parametric estimates of the first pasage density using EULER and saddlepoint inversions, along with the true pdf. Table 3.12 shows integrated absolute errors for the methods. IAE for the parametric saddlepoint and EULER are identical, since both methods estimate the parametric density almost

Method	IAE (nonparametric)	IAE (parametric)
EULER	0.10561	0.27545
Saddlepoint	0.07449	0.27545

Table 3.12: Cumulative damage example: integrated absolute errors (IAE) relative to the true pdf of nonparametric and parametric density estimates.

exactly. However, both estimates, made under the Poisson assumption, are incorrect, and considerably worse than the nonparametric estimates, as shown by comparison of the IAEs.

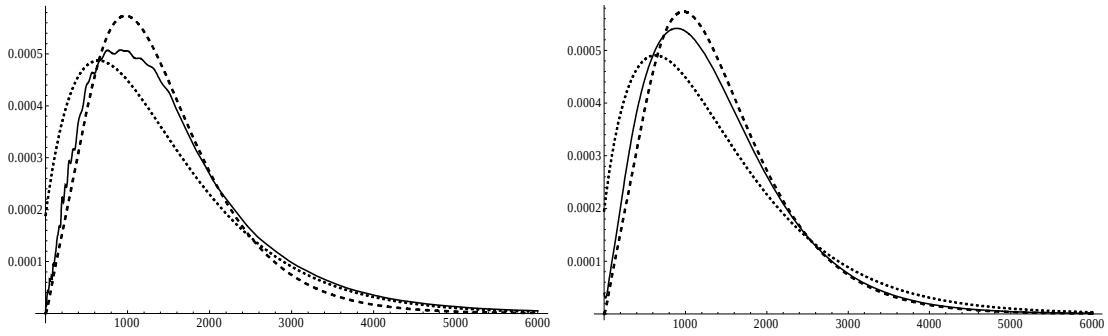


Figure 3.23: Cumulative damage $1 \rightarrow 4$ first passage density estimates. Left: Smoothed empirical EULER (solid), parametric EULER (dotted), true density (dashed). Right: Empirical saddlepoint (solid), parametric saddlepoint (dotted), true density (dashed).

Though it is not the main point here, this example also shows the advantage of using methods that are not restricted to Markov processes, where the assumption of exponential holding times is built into the model. Semi-Markov models are more general, and quite tractable using flowgraph methods.

Chapter 4

Discussion

4.1 Summary of results

In this dissertation we have provided background on statistical flowgraphs and showed how previous work can be significantly extended in the direction of nonparametric and semiparametric flowgraphs.

We began in Chapter 1 by carefully formulating a class of problems that can be readily solved using flowgraph methods, and can only be solved with greater difficulty, or not at all, using other methods. The crux of these problems, with many examples in reliability and survival analysis, is to estimate the distribution of first passage times between two arbitrary states in a multistate system, when only data on transitions between adjacent states are assumed known.

Chapter 1 continued with a review of parametric flowgraphs, where we assume a known family of distributions $F_{ij}(t;\theta_{ij})$, parameterized by θ_{ij}, that determines the conditional holding time in state i given a transition to j. We then reviewed the transforms used in flowgraph analysis at the highest level of generality, namely as

Lebesgue-Stieltjes integrals, while providing a coherent exposition of results scattered in the literature that form a rigorous basis for the flowgraph methodology. In particular we showed the validity of Mason's rule, a general procedure for solving flowgraphs of arbitrary complexity. Since the empirical transforms used in Chapter 2 are also Lebesgue-Stieltjes integrals, this rigorous basis carries over to the nonparametric flowgraph methodology.

Chapter 2 is the core of the dissertation, presenting a set of original results and algorithms, based on limited prior work, that demonstrate the existence of effective and useful techniques for solving flowgraph problems nonparametrically (with no distributional assumptions) or semiparametrically (with assumed models for holding times on some, but not all, of the transitions in a multistate system). After reviewing empirical transforms (integral transforms based on a sample average), the following key results were presented in Chapter 2:

- Showing that empirical transforms for individual transitions combine using Mason's rule to yield an empirical transform for the desired first passage distribution, that it is essentially equivalent to an empirical transform based directly on a sample of first passage times, and that it is a strongly consistent estimator of the exact transform of the true parametric first passage distribution.
- Showing that the same results hold when transition samples are randomly right censored.
- Developing and evaluating in depth several algorithms for numerical inversion of empirical transforms, using Fourier series approximation and the saddlepoint method.
- Proving the nonexistence of any bound on the error of the saddlepoint density approximation; this result is important because, while known heuristically, it is widely ignored in the saddlepoint literature.

- Quantifying bounds for the domain over which the empirical saddlepoint equation can be solved, thus setting upper and lower bounds on the domain for an empirical saddlepoint approximation. These bounds were known for transforms directly based on a single sample, but not for complex transforms computed using Mason's rule.

- Showing the nonexistence of exact confidence bounds for the pdf and CDF of the first passage distribution, using the estimation methods we developed or anything similar; and developing a bootstrap procedure for confidence bounds on moments of the first passage distribution.

- Showing that due to the generality of the results on transforms and Mason's rule, we can freely mix parametric and empirical transforms in a single computation, leading to a flexible semiparametric flowgraph method.

- Developing computer code using Mathematica to implement all the parametric and nonparametric methods described here. (Appendix A to the dissertation contains algorithm descriptions and annotated code listings for programs written to support the research.)

Methods were analyzed in terms of accuracy and computational complexity, and important points were illustrated with examples. We also discussed how the results developed here relate to other work in the field.

Chapter 3 presented three detailed examples, with variants on each, to illustrate our methods in the context of realistic problems in reliability analysis and structural engineering. By using simulated data, we were able to quantify the accuracy of the methods using standard statistics, besides using plots for visual comparison of estimates to the true distributions.

This chapter, besides summarizing, presents some general thoughts on parametric and nonparametric methods in the broader scientific context, leading into a discussion

of where our particular nonparametric methods fit into a general methodology for analyzing stochastic multistate systems. Finally, we list an agenda of future research objectives for the purpose of extending these methods and making them more useful, and easier to use.

4.2 Theory-based versus data-driven models

While probability theory is easily categorized as a purely deductive mathematical discipline (Kolmogorov 1956), statistics, like science in general, is a dynamic balancing act between discovering theory through induction and validating theory through prediction and experiment (Sprenger 2009). Statistics serves science, but can also be taken as a science in itself if we "abstract away" the subject-matter content from statistical theories.

Rather than try to define "science," we take physics as the paradigm of a mature science. As an example of historical progress in physics, consider Snell's law of refraction (Holton & Brush 2001, Section 3.4, 23.1): if θ_1 is the angle of incidence of a light ray at the interface between two different transparent media (e.g., air and glass), it is related to θ_2, the angle of refraction, by $\sin\theta_1/\sin\theta_2 = c$, where c is a constant determined by the two substances. The stages leading to a full understanding of this law are illustrative of progress in science generally.

At the most primitive stage that can be called real science, we see pure empirical ("nonparametric") laws. In the case of refraction, this goes back at least to Claudius Ptolemy, who posited a proportionality between the angles of incidence and refraction (not their sines) based on experimental observations of refraction between air and water; this law was valid only for small angles, and there was no good reason to suppose that the same law would be applicable to other substances. Eventually,

a "parametric" mathematical law was discovered (c being the parameter) independently by several scientists, one of whom was Snell in 1621. This is what statisticians would call a model. The final stage is a parametric law derived from more basic explanatory principles of the science; the explanatory principles constitute what physicists call a model (we will call it a "theory"). In the case of Snell's law, this was done, incorrectly, by both Descartes and Newton, based on a particle theory of light; it was later done "correctly" by Fresnel based on a wave theory of light. The theory is not mere cognitive icing on the cake, but leads to genuine insight—in this case, that the constant c is the ratio of the velocities of light in the two media. "Correctly" appears in quotes because such theories are subject to change, even when the form of the mathematical law remains the same; though Fresnel's explanation is still taught, a different and more "correct" version is provided by the theory of quantum electrodynamics (Feynman 2006).

Using models that are "parametric" in the statistical sense does not necessarily imply a motivation towards theory-based understanding of a problem. One of the authors of a leading textbook on simulation has developed a software package that automates the process of selecting parametric models for data sets, based on maximum likelihood estimation of the parameters and goodness-of-fit testing: "Performing the statistical procedures discussed in this chapter can . . . be difficult, time-consuming, and prone to error . . . ExpertFit[1] will automatically and accurately determine which of 39 probability distributions best represents a data set" (Law & Kelton 2000, p. 370). An instructive example of the merits of this approach is found in (Wolstenholme 1999), Example 3.5; a much-analyzed data set of failure times for ball bearings is found (by Wolstenholme and others) to be a reasonable fit to several distributions commonly used in reliability analysis, e.g., lognormal, gamma, and Weibull. None of these analyses has been based on physical properties of the bearings (i.e., on a theory), and none has contributed anything to a scientific

[1]Trademark of Averill M. Law & Associates

understanding of why ball bearings fail—even though they may have some predictive value for bearings with the same characteristics as those analyzed.

A statistical example more suggestive of theory is the derivation of the Weibull distribution from the weakest-link property, i.e., based on the distribution of the smallest failure times for components in series (Wolstenholme 1995; Mann *et al.* 1974, Section 3.4). A similar example is Daniels' (1945) analysis of cumulative damage to bundles of threads. From the standpoint of our paradigm mature science, these examples are important because they may point the way to correct theoretical understanding of important phenomena. Such results may also generalize to powerful theories in statistical science, as in the case of the central limit theorem, which evolved out of solutions to scientific problems investigated by De Moivre, Laplace, and Gauss (Stigler 1986).

From the point of view of science, nonparametric statistical results (and "parametric" results derived from blindly fitting data against a panoply of distributions) may be steps on the way to theory, and may enable prediction within a limited range, but cannot be ends in themselves. (Here we are not considering statistical methods such as hypothesis testing for the *validation* of theories, but rather statistics as a tool for *discovering* theories.)

Breiman (2001) views natural phenomena as "black boxes" whose inputs and outputs can be observed, and he talks about two statistical cultures: one that looks for parametric models that are "good fits" for the data, another that "considers the inside of the box complex and unknown. Their approach is to find a function $f(\mathbf{x})$—an algorithm that operates on $\mathbf{x}$ to predict the responses $\mathbf{y}$" (*ibid.*, p. 199). Breiman sees accurate prediction as the *raison d'être* of statistics, and certainly this is true as far as it goes. But consider, e.g., perhaps the most widely used tool of prediction, the linear statistical model. Even within the range of observed data, valid use of the model depends on assumptions of linearity and inclusion of all significantly relevant

predictor variables (Christensen 1998, Section 13.5). These assumptions amount to an unstated theory that the underlying causal phenomena are known, and have a linear effect (at least approximately) on the observed response, and that future predictor values will be similar to those used to build the model.

By adding enough predictor variables, we may create a model with good fit to the data at hand, but little ability to predict future data. Disparaging such models, John von Neumann is quoted as saying "with four parameters I can fit an elephant, and with five I can make him wiggle his trunk" (Dyson 2004). Even where such a model is consistently good at prediction, it may not provide any insight into a causal mechanism that explains events. Taking relativity as an example, the physicist David Deutsch says (1998, p. 3)

> What makes the general theory of relativity so important is not that it can predict planetary motions a shade more accurately than Newton's theory can, but that it reveals and explains previously unsuspected aspects of reality, such as the curvature of space and time.

Thus when a model rises to the level of theory, it may enable prediction of phenomena far beyond the range of what the original data suggested.

We would argue for a third statistical culture in addition to Breiman's two: statisticians who believe that developing theories of causal mechanisms underlying their models may result in better predictions, and provide additional insight into the prediction process. This viewpoint is implicit in much of the literature on applied stochastic processes, and has emerged recently as a trend in survival analysis. The following quote, from (Aalen & Gjessing 2001), nicely summarizes the general idea:

> In survival analysis one studies the time to occurrence of some event ... one wishes to analyze the probability distribution of the time to the event

> by means of survival curves and hazard rates. What is, however, usually disregarded in the standard approach ... is that the event is the end point of some process ... the underlying process leading to the event is largely unknown. This, however, does not imply that it should be ignored. Considering the underlying process, even in a speculative way, may improve our understanding of the hazard rate and give alternative regression models.

This view is supported by Künsch (2008) , whose experience with scientists as consumers of statistical models leads him to a list of desiderata for such models. Paraphrasing Künsch, a model should:

- Take as much knowledge of the underlying process as possible into account;
- contribute to understanding the process;
- be applicable to similar processes;
- allow prediction of the process under conditions not yet observed;
- have parameters with a clear interpretation in the scientific domain.

This is a far cry from Breiman's (2001) statistical cultures, where

> ... the models that best emulate nature in terms of predictive accuracy are also the most complex and inscrutable ... The point of a model is to get useful information about the relation between the response and predictor variables ... The goal is not interpretability, but accurate information.

"Accurate information" is a necessary, but not sufficient, criterion for a scientific model, and this opinion leads to our position on appropriate uses of the nonparametric methods developed in this dissertation.

4.3 Uses of nonparametric flowgraph methods

Analysis of a process tending towards some end state such as death or failure of a system is more informative if it takes into account the finer details of the process, as recommended by Aalen and Gjessing. This typically means decomposing the process into a series of intermediate states traversed *en route* to the end state. Ideally, we have, or hope to discover, some basis in theory for choosing parametric probability models for transitions between adjacent states. Queueing theory provides examples, where, e.g., there are often sound reasons for postulating a Poisson arrival process driving state transitions (Kleinrock 1975).

There is a danger in using established statistical theory, since it tends to be codified into simplistic rules of thumb; for example, many reliability engineers automatically assume a Weibull failure distribution unless there is compelling evidence to the contrary. As another example, a well-known methodology used for decades in reliability prediction of complex military electronic systems (DoD 1995) assumes exponential failure distributions, mostly on account of mathematical tractability. (This is comparable to the over-assumption of normality in other branches of statistics.) Where no theoretical basis exists, the typical advice to practicioners is to select models based on visual inspection of a histogram or other density estimate derived from sample data (Huzurbazar 2005a; Law & Kelton 2000, Chapter 6). The discussion in the previous section regarding the ball bearing failure data shows the risk in either of these procedures. Typically more than one model will fit the data, and competing models can only be discriminated based on information, such as tail behavior, that cannot be inferred from the sample unless it is very large.

This problem becomes more significant in multistate models with many transitions, where there are many models to be hypothesized and incorrect choices can cascade in estimates for first passage times between the states of interest. Where

evidence for particular parametric families is not compelling (and it rarely is), nonparametric methods are appropriate as an adjunct to the use of parametric models. Unlike Breiman (2001), we do not believe that nonparametric methods can *supplant* parametric models, either in general or in any particular problem. We feel that the goal, reflecting the collaboration between statistics and other sciences, should always be a theoretically grounded parametric model.

In the present context of multistate flowgraph models, the methods described in this dissertation can advance us towards that goal by supplementing parametric flowgraphs in various ways:

- Where the evidence for a particular parametric family is not compelling, where there is controversy over the "correct" model to be used, or where for any other reason the researcher wishes to minimize *a priori* assumptions, the methods described here will lead to a result. The error of the result is limited by the data, assuming the use of both saddlepoint and Fourier series methods to validate the first passage transform inversion. We can then hope that replications of the sampling process will yield information on the stability of the nonparametric result, and ideally will lead to a better understanding of the process as a whole, ultimately contributing to a theoretically grounded understanding.

- Where the researcher feels that the hypotheses implied by a parametric model are justified, The methods described here are valuable for comparison and validation—if the nonparametric result is confirmatory, it strengthens the case for a particular parametric model; if not, it alerts the researcher that further analysis or more data are needed.

- Studying a particular stochastic model through many examples may lead to a gradual unfolding of insight into the mechanisms underlying the various transitions. Using the semiparametric approach described in Section 2.4, transition

mechanisms can be incorporated into the model as parametric distributions in a stepwise fashion, leaving the less well understood transition distributions in nonparametric form. Each step (if justified) increases the accuracy of the results.

- Though the nonparametric results are not very informative for small samples, they are honest in the sense of relying only on the data, whereas basing a parametric model on small samples can lead to unjustified confidence in a model with little real support. A false level of certainty may inhibit further research and delay obtaining a better theoretical understanding of the process.

Based on the results reported in this dissertation, nonparametric methods for statistical flowgraphs are valuable adjuncts to existing parametric methods. Further research into issues raised by this study, which may make these methods more useful, is described in the next section.

4.4 Future directions

- *Accuracy of the saddlepoint density approximation:* As discussed in Section 2.2.3, Daniels (1980) showed that the saddlepoint density approximation is exact for certain families of distribution; on the other hand, Theorem 2.2.1 demonstrates that, in general, the error in the approximation may be arbitrarily large. The usefulness of the saddlepoint method for densities would be increased if we could predict, from characteristics of the MGF or CGF, some bound on the possible error. This might also allow construction of confidence bounds for points of the approximation.

- *Use of Bayesian methods:* Scientists and engineers often have prior beliefs about transition probabilities and holding time distributions in multistate ap-

plications; incorporating expert opinion into nonparametric flowgraph modeling, using methods such as those mentioned in Section 2.2.4, may prove fruitful.

- *Confidence bounds:* Pointwise confidence intervals or confidence bands for the first passage distribution function would be very desirable. Though it is not clear how to improve the current situation (see Section 2.3), the importance of the problem makes it a topic for future research. A related problem is to extend Theorem 2.1.2 to show that a Mason's rule expression based on empirical transforms is unbiased for finite samples, as well as consistent.

- *Software improvements:* The algorithms and coding need to be improved in two areas: performance and ease of use. Currently, producing the EULER and saddlepoint approximations takes minutes, even for fairly simple flowgraphs. To be usable for a range of practical problems, this needs to be speeded up. In addition, solving a problem now is done with Mathematica code that needs a significant amount of customization for the particular flowgraph being analyzed. Ideally, we would like to generalize all the functions to the point where a user could enter the problem data, push a botton, and see the results.

Appendix A

Appendix: Mathematica code

Mathematica (Wolfram 2003; Hilbe 2006) was used for all numeric and symbolic computations done in support of this dissertation. Mathematica is typically regarded as "computer algebra" software, but also has powerful numeric capabilities and a very large library of functions for general mathematics, probability and statistics. This appendix contains algorithm descriptions and annotated code listings for programs written to support the research. Routine coding to produce graphic plots is not included.

Mathematica supports various programming styles, including functional and procedural, and can also be written in a style resembling conventional mathematical notation. Mathematica comments are enclosed by `(* ... *)`. In this document computer code and output is shown in a typewriter font:

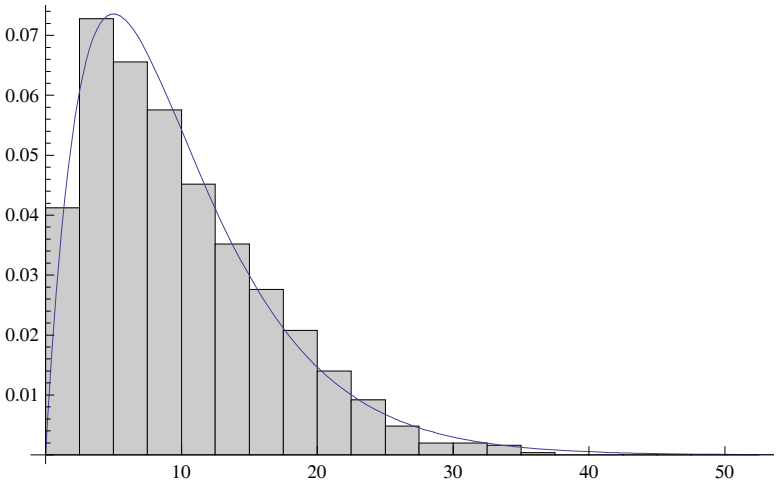
```
(* This is Mathematica code *)
y = Sum[(x/2)^2, {x, 1, 10}]
(* Or it can be written in a more math-like format *)
```

$\mathtt{y} = \sum_{\mathtt{x}=1}^{10} \left(\frac{\mathtt{x}}{2}\right)^2$

The next page shows an example of an interactive Mathematica session. A sample is generated from a gamma distribution, the empirical transform is computed, and the EULER inversion is plotted against the true density.

In[39]:=
```
(* Define gamma(2,5) distribution, generate 1,000 sample points *)
gammaDist = GammaDistribution[2., 5.];
gammaPdf[t_] = PDF[gammaDist, t];
gammaSamp = RandomReal[gammaDist, 1000];
(* Plot the sample histogram against the density *)
Show[
 Histogram[gammaSamp, HistogramScale → 1,
  BarStyle → GrayLevel[0.8], HistogramCategories → 20, HistogramRange → {0, 50}],
 Plot[gammaPdf[t], {t, 0,50  }]]
```

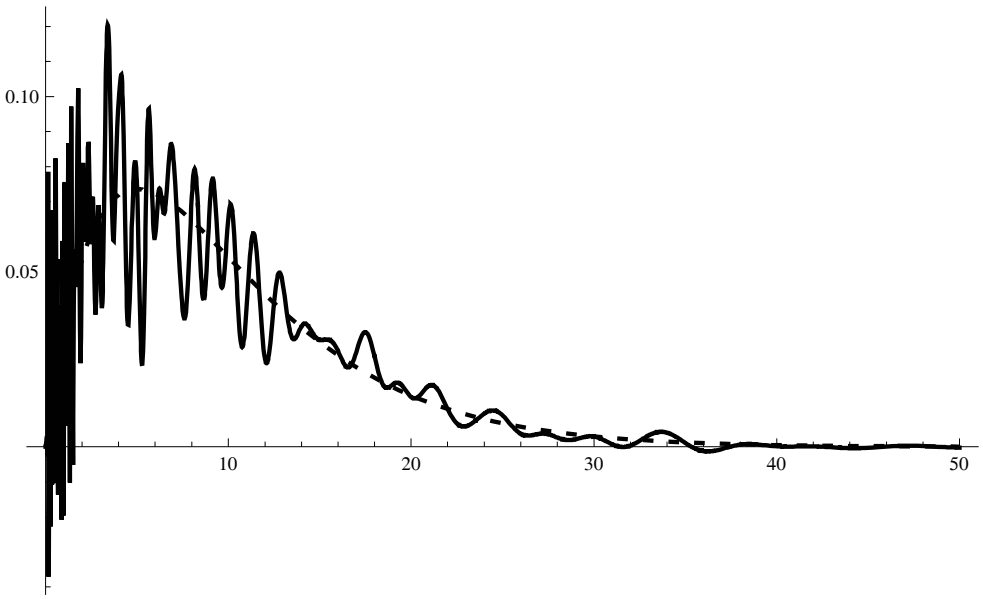
Out[42]=



In[43]:=
```
(* Compile the computation of the empirical Laplace transform *)
ceGammaL = Compile[{{s, _Complex}}, (emplt[gammaSamp])[s]];
(* Plot EULER inversion against true density *)
ψ[s_] := Re[ceGammaL[s]];
```

In[45]:=
```
Plot[{euler[ψ, t], gammaPdf[t]}, {t, 0, 50}, PlotRange → All,
 PlotStyle → {{Black, AbsoluteThickness[2]}, {Black, AbsoluteThickness[2], Dashing[Medium]}}]
```

Out[45]=



A.1 General functions

We describe here some general functions used in various places.

A.1.1 Function compilation

Mathematica is normally an interpreted language. In some cases, functions can be compiled into machine code, which reduces the execution time by about a factor of 10. In order to be compilable functions must conform to certain restrictions, such as taking only simple arguments (real, complex, integer, boolean, or arrays of simple arguments), and using machine precision (in general, Mathematica allows arbitrary precision). Here is an example of normal function definition followed by a compiled function definition:

```
uniPdfLT[s_] := 1 − e^−hs/(hs);
uniPdfLT := Compile[s, _Complex, 1 − e^−hs/(hs)];
```

Where possible, we use compiled functions for reduced overhead; however in the code listings in this appendix, for simplicity we usually do not show the compilation.

A.1.2 Mathematica's `InterpolatingFunction`

The Mathematica function `Interpolation` takes a list of (x, y) coordinates and returns an `InterpolatingFunction` object, a continuous function that interpolates between the values of the given points (or extrapolates for values outside the range of the points). The interpolation is linear, quadratic, etc. depending on the value of the parameter `InterpolationOrder`. The default cubic interpolation is highly accurate for smooth functions (e.g., probability densities) when the set of points on which the interpolation is based is sufficiently dense. For inversions based on sam-

ples, we use linear interpolation, which preserves monotonicity; this is important for CDF approximations, and also for densities, which are locally monotonic. Polynomial interpolation may produce small bumps in the approximations. In practice, an interpolation of 500 points computed over the effective support of a smooth density is sufficient to provide an IAE on the order of 10^{-5} (assuming the points represent the exact values—for approximations, we use a finer mesh of points).

Our main motivation for using these functions is that producing evaluation points of the inversion algorithms is costly, and interpolating functions allow sets of points to be efficiently reused for plotting and numeric integration.

The following example shows that approximating a gamma density with 500 points over the range [0, 60] produces a very small error, as measured by IAE:

```
gamPdf[t_] := PDF[GammaDistribution[2, 5], t];
Δ = 60/500.;
gamApprox = Interpolation[Table[{t, gamPdf[t]}, {t, 0, 60, Δ}],
   InterpolationOrder -> 1];
NIntegrate[Abs[gamApprox[t] - gamPdf[t]], {t, 0, 60}]
0.0000609578
```

(With the default cubic interpolation, the IAE in this example would be 1.58294×10^{-8}.)

Other examples appear in the the following sections, e.g., in the `estcdf` function in the next section.

A.1.3 Estimating the distribution function

The function `estCdfPoints` takes a set of points of the form $(t, \hat{f}(t))$, where $\hat{f}$ is a density estimate, and returns a set of points $(t, \hat{F}(t))$, where $\hat{F}$ is an estimate

of the CDF computed by trapezoidal-rule integration. The function `estCdf` takes the same set of points as input and returns an interpolating function, which uses linear interpolation (to insure that the approximate CDF is monotonic) to provide a continuous approximation to the CDF.

```
estCdfPoints[points_] := Module[
   {Δx, summands, ySum},
   Δx = Differences[Table[points[[i,1]], {i,1,Length[points]}]];
   summands =
      Table[((points[[i,2]] + points[[i + 1,2]])/2)Δx[[i]],
        {i,1,Length[points] - 1}];
   ySum = Accumulate[summands];
   Table[points[[i,1]], ySum[[i]], {i,1,Length[points] - 1}]
];
estCdf[points_] :=
   Interpolation[estCdfPoints[points],InterpolationOrder -> 1];
```

For plotting the CDF or calculating the IAE relative to the true CDF we use `estCdf`; `estCdfPoints` is used to calculate the Kolmogorov-Smirnov statistic. The K-S statistic, commonly used to assess the fit of two distribution functions, is the maximum absolute difference between the CDFs. The function `statisticKS` returns the Kolmogorov-Smirnov statistic, given a set of (x, y) points representing a density approximation, and a closed-form CDF function.

```
statisticKS[points_, cdf_] := Module[
   {eCdfPoints},
   eCdfPoints = estCdfPoints[points];
   Max[
      Table[Abs[eCdfPoints[[i, 2]] - cdf[eCdfPoints[[i, 1]]]],
        i, 1, Length[eCdfPoints]]
```

```
   ]
]
```

A.1.4 Root finding

Solving the empirical saddlepoint equation (see Section 2.2.3) requires finding a zero of $\tilde{K}'(\hat{s}) - t$, where $\tilde{K}$ is the empirical CGF. Because of the complexity of the ECGF we have found that conventional root-finding algorithms such as Newton-Raphson and others used by Mathematica's `FindRoot` are not very robust. To avoid having to do extensive error checking, we use bisection to solve the saddlepoint equation. This is implemented in the function `bisection`; `f` is the function for which a zero is to be found, `(x0, x1)` is the starting interval, `maxIter` is the maximum number of bisections to perform, and `tolerance` is used to determine when a solution has been found.

```
bisection[f_, x0_, x1_, maxIter_, tolerance_] := Module[
   {mid, val, k = 0, a = N[x0], b = N[x1]},
      mid = (a + b)/2;
      val = f[mid];
      While[(k < maxIter && Abs[val] > tolerance),
        If[ Sign[f[a]] == Sign[f[mid]], a = mid, b = mid];
        mid = (a + b)/2;
        val = f[mid];
        k = k + 1];
      mid
]
```

A.2 Computing exact and empirical transforms

For exact transforms, Mathematica has a variety of functions relating to Laplace and Fourier transforms. We illustrate using Laplace transforms. The MGF is handled by simply reversing the sign of the argument to the Laplace transform.

This example defines a gamma pdf, takes its Laplace transform, then inverts the transform to recover the pdf (Mathematica outputs are indented):

```
gammaDist = GammaDistribution[3, 5];
gammaPdf[t_] = PDF[gammaDist, t]
```

$$\frac{1}{250} e^{-t/5} t^2$$

```
L[s] = LaplaceTransform[gammaPdf[t], t, s]
```

$$\frac{1}{(1+5s)^3}$$

```
InverseLaplaceTransform[L[s], s, t]
```

$$\frac{1}{250} e^{-t/5} t^2$$

Mathematica's inversion capability is quite powerful, certainly beyond what one could easily accomplish manually. Exact inversion of empirical transforms is possible in principle, but computationally intractable for practical problems.

The function `emplt` implements the empirical Laplace transform; `samp` is the basis sample, `&` at the end of the definition indicates that an anonymous function is being returned, and `#1` is the argument to the anonymous function.

```
emplt[samp_] := Total[e^-samp#1]/Length[samp]&;
```

Since this function involves only simple arguments (complex numbers) we compile it for use, e.g., where `gammaSamp` is a basis sample, this code produces the compiled empirical transform `ceGammaL`, taking one argument $s \in \mathbb{C}$:

```
ceGammaL = Compile[{{s, _Complex}}, (emplt[gammaSamp])[s]];
```

A.2.1 Empirical transforms of censored data

The function `simulateCensoredData` is used to generate simulated censored data. Given an uncensored sample of failure times and a censoring distribution, it returns a table with the minimum of (censoring time, failure time) plus a censoring indicator. The returned value is a list of pairs (minimum of the two times, censoring indicator). The censoring indicator is 1 = uncensored observation, 0 = censored observation. The list is sorted by failure or censoring time.

```
simulateCensoredData[cDist_, fSample_] := Module[
   {cens, ret, n},
   n = Length[fSample];
   cens = Round[RandomReal[cDist, n]]; (* Generate censoring times *)
   ret =
      Table[Min[cens[[i]], fSample[[i]]],
         If[fSample[[i]] <= cens[[i]], 1, 0], i, 1, n];
      Sort[ret, #1[[1]] <= #2[[1]]&]
];
```

Typically we use an exponential(λ) distribution for the censoring distribution, and adjust λ to provide the desired percentage of censoring (which is determined by the proportion of 0s in the output from `simulateCensoredData`).

The function `empltCensored` implements the empirical Laplace transform for a censored sample. Its argument `samp` has pairs (t, c) where t is a sample time and c is 1 if t is a failure time, 0 if it is a censoring time. Efron's "redistribute to the right" algorithm is used to weight the points (see Section 2.1.4). It returns an anonymous function taking one argument, like `emplt`.

```
empltCensored[samp_] := Module[
   {n = Length[samp], initialMass, probs, massToRedistribute, points,
```

```
   portion, empiricalMass},
   initialMass = N[1/n];
   probs = Table[initialMass, i, 1, n];
   points = Table[0, 0, i, 1, n];
   Do[(
      If[samp[[i]][[2]] == 0,
        (massToRedistribute = probs[[i]];
        probs[[i]] = 0;),
        massToRedistribute = 0
      ];
      portion = massToRedistribute/(n - i);
      Do[probs[[j]] = probs[[j]] + portion, j, i + 1, n];
      ), i, 1, n - 1
   ];
   Do[(
      If[probs[[i]] == 0, ,
        points[[i]] = probs[[i]], samp[[i]][[1]]
      ];
      ), i, 1, n
   ];
   empiricalMass = Select[points, #[[1]] != 0 &];
   Sum[empiricalMass[[i]][[1]] Exp[-# * empiricalMass[[i,2]]],
      i, 1, Length[empiricalMass]] &
];
```

Code for inversion of empirical transforms is presented in the following sections.

A.3 Fourier series approximation

This section presents Mathematica code for implementing the methods described in Section 2.2.2.

A.3.1 The EULER algorithm

The function `euler` implements Abate and Whitt's EULER algorithm. The Mathematica code is a fairly direct translation from the UBASIC code given on pages 7-8 of (Abate & Whitt 1992), with roughly the same variable names. The last line is the Euler summation in Equation (2.25) (p. 81). the function takes as arguments a function `fn` which implements $\mathrm{Re}[\mathcal{L}(f)]$, the real part of the Laplace transform to be inverted, and the point t at which the inversion is to be computed; it returns $f(t)$.

```
euler[fn_,t_]:= Module[
   {Su,y,x,h,U,A=18.4,Ntr=15,
     C=Table[Binomial[11,i],{i,0,11}]}
   x = A/(2t); h = π/t; U = e^(A/2)/t;
   Su[1] = fn[x]/2 + Sum[(y = nh; (-1)^n fn[x + iy]), {n, 1, Ntr}];
   Do[
      y = (Ntr + k)h;
      Su[k + 1] = Su[k] + (-1)^(Ntr+k) fn[x + iy],
      {k,12}
   ];
   U Sum[C[[j]]Su[j + 1], {j, 1, 12}]/2048
]
```

This is an example of an input `fn` to `euler`:

```
ceGammaL = Compile[{{s, _Complex}}, (emplt[gammaSamp])[s]];
ψ[s_] := Re[ceGammaL[s]];
```

The `eulerNew` function is a modification of Euler. The binomial table is taken outside the module definition to avoid recalculating it every time the function is called, the number of terms in the partial Fourier sum is reduced, and the location of the contour of integration is shifted towards the origin by changing the variable `A`. In terms of Equation (2.25), we reduce N from 15 to 0 and increase M from 11 to 12. This reduces the number of terms actually computed from 27 to 14, which smooths the result and reduces the computation time. The modifications were heuristically derived.

```
$C=Table[Binomial[12,i],{i,0,12}];
eulerNew[fn_,t_]:= Module[
   {A=4,n=13,Su,y,x,h,U},
   x = A/(2t); h = π/t; U = e^(A/2)/t;
   Su[1] = fn[x]/2;
   Do[
      y=kh; Su[k+1]=Su[k]+(-1)^k fn[x+i y],
      {k,n}
   ];
   U Σ_{j=1}^{n} $C[[j]] Su[j+1] / 2^(n-1)
]
```

A.3.2 Kernel presmoothing

First we provide a little more detail on the kernel presmoothing discussed in Section 2.2.2 (p. 82). To avoid bias, a kernel should distribute the mass of each sample

point symmetrically. Figure A.1 shows an example for a symmetrical rectangular kernel, with the sample points indicated by "×." For a bandwidth h, this kernel is a uniform$(-h/2, h/2)$ density.

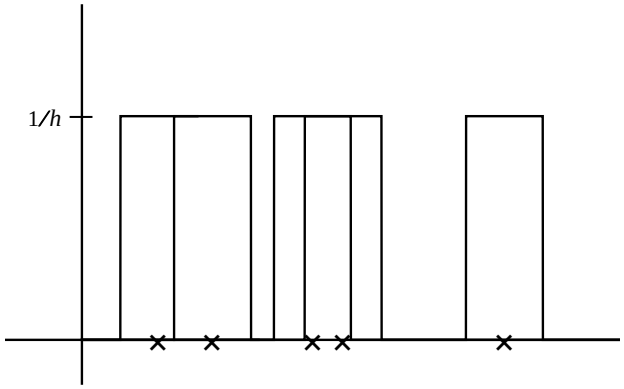


Figure A.1: Symmetric rectangular kernels over sample points

It is obvious from the figure that shifting all the sample points left by $h/2$, then using a uniform$(0, h)$ density as a kernel is equivalent, and computationally more convenient. However, either shifting or using a symmetric kernel will put probability mass to the left of the origin for sample points in $[0, h/2)$. If the density estimate is truncated at 0, this results in bias near the origin. We compensate by reflecting the negative points in the ordinate after shifting, thus moving the mass back to its approximately correct location. This is effected by simply taking the absolute value of the shifted sample points.

One might think that a smoother kernel would produce superior results, but our test cases indicate that the rectangular kernel works just as well as an exponential or gamma kernel, as measured by IAE. This is consistent with the literature on density estimation; e.g., Table 3.1 in (Silverman 1986) shows that the efficiency of the rectangular kernel is within 10% of the optimal Epanechnikov kernel (where efficiency measures the relative sample size needed for the same accuracy).

This is the code used to generate the presmoothed estimate for the small sample from the gamma(2,5) distribution shown in Figure 2.11 (p. 84). The sample points are in `gammaSampSmall`. Here the rectangular kernel function is developed for conventional kernel smoothing:

```
h = 1.5; (* Bandwidth *)
uni = UniformDistribution[{0, h}];
uniPdf[t_] := PDF[uni, t];
gammaSampSmallShifted = Abs[gammaSampSmall - h/2];
nS = Length[gammaSampSmallShifted];
uniKerSmall[x_] := (1/(nS*h))*
   Sum[uniPdf[(x - gammaSampSmallShifted[[i]])/h], {i, 1, nS}];
```

This is used only to produce a plot showing the kernel estimate (e.g., the plots on the left in Figure 2.11).

Here the smoothed empirical transform function is defined; `samp` is the basis sample, `&` at the end of the definition indicates that an anonymous function is being returned, and `#1` is the argument to the anonymous function.

```
uniPdfLT[s] := (1 - e^-hs/(hs); (* Laplace transform of uniPdf *)
smoothedEmplt[samp_] :=
   uniPdfLT[#1](Total[e^-samp#1]/ Length[samp])&;
```

The smoothed empirical transform then becomes the input to EULER, either directly or after being combined with other transforms via Mason's rule.

A.3.3 Exponential smoothing

The function `expSmoothed` implements exponential smoothing as described on page 84, using the Mathematica function `ExponentialMovingAverage` in the forward and

reverse directions. The inputs to `expSmoothed` are a list of plot points produced by EULER, and the smoothing parameter α.

```
expSmoothed[points_, α_] := Module[
   {n, xTable, yTable, left, right},
   n = Length[points];
   xTable = Table[points[[i, 1]], {i, 1, n}];
   yTable = Table[points[[i, 2]], {i, 1, n}];
   left = ExponentialMovingAverage[yTable, α];
   right = Reverse[ExponentialMovingAverage[Reverse[yTable], α]];
   Interpolation[
      Table[xTable[[i]], (left[[i]] + right[[i]])/2, i, 1, n],
      InterpolationOrder -> 1]
]
```

The output of `expSmoothed` is an interpolating function (see Section A.1.2) which is a smoother approximation of the EULER output.

A.4 Saddlepoint approximation

First, here is the computation for the exact (parametric) saddlepoint approximation of a gamma(2,5) density. The pdf is `gammaPdf`, and the subsequent lines define the LT, CGF, and derivatives of the CGF:

```
L[s_] = LaplaceTransform[gammaPdf[t], t, s];
M[s_] = L[-s]; (* Moment generating function *)
K[s_] = Log[M[s]]; (* Cumulant generating function *)
dK[s_] = FullSimplify[K'[s]]; (* 1st derivative of CGF *)
ddK[s_] = FullSimplify[K''[s]]; (* 2nd derivative of CGF *)
```

```
dddK[s_] = FullSimplify[K'''[s]]; (* 3rd derivative of CGF *)
ddddK[s_] = FullSimplify[K''''[s]]; (* 4th derivative of CGF *)
```

`FullSimplify` applies a wide range of algebraic and other transformations to produce the simplest expression of the derivatives.

The function `saddleInv` computes the first-order saddlepoint density approximation for its argument:

```
saddleInv[t_] := (
   sp = Bisection[(dK[#] - t)&, −10 .4, 100, 10^−5];
   (1./Sqrt[2*Pi*ddK[sp]]) Exp[K[sp] - sp*t]
)
```

The first line of the function finds the saddlepoint $\hat{s}$ (see Section 2.2.3); the second line implements the actual approximation (Equation 2.26, p. 92).

The function `saddleInvSecondOrder` computes the second-order saddlepoint density approximation for its argument (Equation 2.29, p. 94). Finding the saddlepoint is the same as in the first-order approximation:

```
saddleInvSecondOrder[t_]:= (
sp = Bisection[(dK[#] - t)&, −10 .4, 100, 10^−8];
firstOrder = Exp[K[sp]- sp*t]/√(2π ddK[sp]);
secondOrder = 1 + ddddK[sp]/(8 ddK[sp]^2) − (5dddK[sp]^2)/(24ddK[sp]^3);
firstOrder * secondOrder)
```

The empirical saddlepoint computation is essentially the same. First the empirical CGF and its derivatives are computed:

```
eK[s_] = Log[ceGammaL[-s]]; (* Empirical CGF *)
edK[s_] = FullSimplify[eK'[s]];
```

```
eddK[s_] = FullSimplify[eK''[s]];
```

The first-order saddlepoint computation is then

```
saddleInv[t_] := (
   sp = Bisection[(edK[#] - t)&, −10 .4, 100, 10^−5];
   (1./Sqrt[2*Pi*eddK[sp]]) Exp[K[sp] - sp*t]
)
```

which is the same as the parametric approximation except for substitution of the empirical CGF. We do not use the second-order empirical saddlepoint because of the complexity of the third and fourth derivatives of empirical transforms, and because there is no great increase in accuracy for cases where the saddlepoint method is effective.

A.5 Bootstrap confidence intervals

Section 2.3 describes a method for obtaining confidence intervals for moments by bootstrap resampling and recomputing the MGF and its derivatives. We illustrate by an example, computing confidence intervals for the mean and standard deviation of the $1 \to 3$ first passage distribution in the repairable redundant system (Section 3.1).

This function takes a set of points as argument and returns a sample (with replacement) of the same length:

```
resample[samp_] := RandomChoice[samp, Length[samp]];
```

This code resamples each basis sample n times (1,000 in this example), recomputes the EMGF, and computes the mean and standard deviation. The collection of

n means and standard deviations is then used to derive the confidence interval, 95% in this case. The function `empMGF` returns the EMGF for the first passage of interest, given a set of basis samples; it is specific to the particular problem.

```
n = 1 000;
means = Table[Null, n]; standardDeviations = Table[Null, n];
Do[
   (s01 = resample[samp01]; s02 = resample[samp02];
   s10 = resample[samp10]; s12 = resample[samp12];
   eMGF[s_] = empMGF[s01, s02, s10, s12, s];
   means[[i]] = eMGF′[0];
   standardDeviations[[i]] = Sqrt[eMGF″[0] - means[[i]]^2]),
   {i, n}
]
{Quantile[means, .05], Quantile[means, .95]}
{Quantile[standardDeviations, .05], Quantile[standardDeviations, .95]}
```

For 1,000 resamples, the calculation takes about 1 minute on a 700 MHz processor.

Glossary

If applicable, each entry is followed by a parenthesized page number indicating where it is defined or first used.

CDF	Cumulative distribution function, same as DF (12).
CF	Characteristic function; the Fourier transform of a probability density. (11).
CGF	Cumulant generating function; the log of the MGF (92).
DF	Distribution function (12).
EDF	Empirical (or sample) distribution function (7).
ELT	Empirical Laplace transform (37).
EMF	Empirical mass function (64).
EMGF	Empirical moment generating function (37).
Exact transform	A transform based on a member of a parametric family of distribution or density functions (36).
$\mathbf{G}(\tau)$	Matrix of first passage distributions $G_{ij}(\tau)$ (5).
I	Indicator function; where S is the specification of a set, $I_S(x) = 1$ if $x \in S$, 0 if not.

iid — Independent and identically distributed.

Kolmogorov-Smirnov statistic — The maximum absolute difference between two distribution functions.

LT — Laplace transform (11).

MGF — Moment generating function (11).

NPMLE — Nonparametric maximum likelihood estimate (107).

Parametric transform — Exact transform.

PL — Product limit estimator (63).

$\mathbf{Q}(t)$ — Semi-Markov kernel (4).

RV — Random variable (unless otherwise stated, assumed to be non-negative).

τ — Calendar time, the time elapsed since a stochastic process was started (4).

$\mathfrak{T}(s)$ — A general integral transform (12).

$\mathfrak{T}(s)$ — A general transmittance (13).

References

Aalen, O. O. and H. K. Gjessing (2001), "Understanding the shape of the hazard rate: a process point of view," *Statistical Science* **16**, 1–22.

Abate, J., and W. Whitt (1992), "The Fourier-series method for inverting transforms of probability distributions," *Queueing Systems* **10**, 5–88.

Abate, J., and W. Whitt (1995), "Numerical inversion of Laplace transforms of probability distributions," *ORSA Journal on Computing* **7**, 36–43.

Andersen, P. (2002), "Multi-state models for event history analysis," *Statistical Methods in Medical Research* **11**, 91–115.

Asmussen, S. (2003), *Applied Probability and Queues*, 2nd Edition, New York: Springer-Verlag.

Bauer, H. (2001), *Measure and Integration Theory*, Berlin: Walter de Gruyter.

Bedford, T., and R. Cooke (2001), *Probabilistic Risk Analysis: Foundations and Methods*, Cambridge: Cambridge University Press.

Bellman, R. E., R. E. Kalaba, and J. Lockett (1966), *Numerical Inversion of the Laplace Transform*, New York: American Elsevier.

Billingsley, P. (1979), *Probability and Measure*, New York: John Wiley & Sons.

Billinton, R. and R. N. Allan (1992), *Reliability Evaluation of Engineering Systems: Concepts and Techniques*, New York: Plenum Press.

Bogdanoff, J. L. and F. Kozin (1985), *Probabilistic Models of Cumulative Damage*, New York: John Wiley & Sons.

Booth, J.G., and A. T. A. Wood (1995), "An example in which the Lugannani-Rice saddlepoint formula fails," *Statistics and Probability Letters* **23**, 53–61.

Breiman, L. (2001), "Statistical modeling: the two cultures," *Statistical Science* **16**, 199–215.

Briand, D., and A. V. Huzurbazar (2008), "Bayesian reliability applications of a combined lifecycle failure distribution," *Journal of Risk and Reliability* **222**, 713–720.

Burk, F. E. (2007), *A Garden of Integrals*, Washington, DC: Mathematical Association of America.

Butler, R. W. (2000), "Reliabilities for feedback systems and their saddlepoint approximation," *Statistical Science* **15**, 279–298.

Butler, R. W. (2001),"First passage distributions in semi-Markov processes and their saddlepoint approximation," in Saleh, A. K. (ed.), *Data Analysis from Statistical Foundations*, Hauppauge, NY: Nova Science Publishers, 347–368.

Butler, R. W. (2007), *Saddlepoint Approximations with Applications*, Cambridge: Cambridge University Press.

Butler, R. W., and D. A. Bronson (2002), "Bootstrapping survival times in stochastic systems by using saddlepoint approximations," *Journal of the Royal Statistical Society: Series B* **64**, 31–49.

Butler, R. W., and A. V. Huzurbazar (2000), "Bayesian prediction of waiting times in stochastic models," *The Canadian Journal of Statistics* **28**, 311–325.

Casella, G. and R. L. Berger (2002), *Statistical Inference*, 2nd Edition, Pacific Grove, CA: Duxbury.

Chatfield, C., A. B. Koehler, J. K. Ord, and R. D. Snyder (2001), "A new look at models for exponential smoothing," *The Statistician* **50**, 147–159.

Chatfield, C. (2004), *The Analysis of Time Series*, Boca Raton, FL: Chapman & Hall/CRC.

Chen, S. X. (2000), "Probability density estimation using gamma kernels," *Annals of the Institute of Statistical Mathematics* **52**, 471–480.

Chen, W-K. (1967), "On directed graph solutions of linear algebraic equations," *SIAM Review* **9**, 692–707.

Christensen, R. (1998), *Analysis of Variance, Design and Regression*, Boca Raton, FL: Chapman & Hall/CRC.

Christensen, R., T. Hanson, and A. Jara (2008), "Parametric nonparametric statistics: an introduction to mixtures of finite Polya trees," *The American Statistician* **62**, 296–306.

Chu, C.-K., and J. S. Marron (1991), "Choosing a kernel regression estimator" (with discussion), *Statistical Science* **6**, 404–436.

Chung, K. L. (2001) *A Course in Probability Theory*, 3rd Edition, San Diego, CA: Academic Press.

Çinlar, E. (1975), *Introduction to Stochastic Processes*, Engelwood Cliffs, NJ: Prentice-Hall.

Collins, D. H., and A. V. Huzurbazar (2008), "System reliability and safety assessment using nonparametric flowgraph models," *Journal of Risk and Reliability* **222**, 667–674.

Copson, E. T. (1965), *Asymptotic Expansions*, Cambridge: Cambridge University Press.

Cramèr, H. (1946), *Mathematical Methods of Statistics*, Princeton, NJ: Princeton University Press.

Crowder, M. (2001), *Classical Competing Risks*, Boca Raton, FL: Chapman & Hall/CRC.

Csenki, A. (2008), "Flowgraph models in reliability and finite automata," *IEEE Transactions on Reliability* **57**, 355–359.

Csörgő, S. (1982), "The empirical moment generating function," in Gnedenko, B. V., M. L. Puri, and I. Vincze (eds.), *Nonparametric Statistical Inference*, Amsterdam: North-Holland, 139–150.

Csörgő, S., and J. L. Teugels (1990), "Empirical Laplace transform and approximation of compound distributions," *Journal of Applied Probability* **27**, 88–110.

Daniels, H. E. (1945), "The Statistical theory of the strength of bundles of threads," *Proceedings of the Royal Society of London: Series A* **183**, 405–435.

Daniels, H. E. (1954), "Saddlepoint approximations in statistics," *The Annals of Mathematical Statistics* **25**, 631–650.

Daniels, H. E. (1980), "Exact saddlepoint approximations," *Biometrika* **67**, 59–63.

Daniels, H. E. (1982), "The saddlepoint approximation for a general birth process," *Journal of Applied Probability* **19**, 20–28.

Davies, B., and B. Martin (1979), "Numerical inversion of the Laplace transform: a survey and comparison of methods," *Journal of Computational Physics* **33**, 1–32.

Davison, A. C., and D. V. Hinkley (1988), "Saddlepoint approximations in resampling methods," *Biometrika* **75**, 417–431.

Davison, A. C., and D. V. Hinkley (1997), *Bootstrap Methods and Their Application*, Cambridge: Cambridge University Press.

Deutsch, D. (1998), *The Fabric of Reality*, New York: Penguin Books.

Devroye, L., and L. Györfi (1985), *Nonparametric Density Estimation: The L_1 View*, New York: John Wiley & Sons.

Devroye, L., and G. Lugosi (2001), *Combinatorial Methods in Density Estimation*, New York: Springer-Verlag.

DoD (Department of Defense) (revised 1995), *MIL-HDBK-217F: Reliability Prediction of Electronic Equipment*, Washington, DC: United States Department of Defense.

Doetsch, G. (1974), *Introduction to the Theory and Application of the Laplace Transform*, Berlin: Springer-Verlag.

Dubner, H., and J. Abate (1968), "Numerical inversion of Laplace transforms by relating them to the finite Fourier cosine transform," *Journal of the ACM* **15**, 115–123.

Duffy, D. G. (1993), "On the numerical inversion of Laplace transforms: comparison of three new methods on characteristic problems from applications," *ACM Transactions on Mathematical Software* **19**, 333–359.

Dyson, F. (2004), "A meeting with Enrico Fermi," *Nature* **427**, 297.

Efromovich, S. (1999), *Nonparametric Curve Estimation*, New York: Springer-Verlag.

Efron, B. (1967), "The two sample problem with censored data," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA: University of California Press, 831–853.

Elkins, D. A., and M. A. Wortman (2002), "On numerical solution of the Markov renewal equation: tight upper and lower kernel bounds," *Methodology And Computing In Applied Probability* **3**, 239–253.

Escobar, M. D., and M. West (1995), "Bayesian density estimation and inference using mixtures," *Journal of the American Statistical Association* **90**, 577–588.

Feller, W. (1968), *An Introduction to Probability Theory and Its Applications, Volume I*, 3rd Edition, New York: John Wiley & Sons.

Feller, W. (1971), *An Introduction to Probability Theory and Its Applications, Volume II*, 2nd Edition, New York: John Wiley & Sons.

Ferguson, T. S. (1973), "A Bayesian analysis of some nonparametric problems," *The Annals of Statistics* **1**, 209–230.

Feuerverger, A. (1989), "On the empirical saddlepoint approximation," *Biometrika* **76**, 457–464.

Feuerverger, A., and R. A. Mureika (1977), "The empirical characteristic function and its applications," *The Annals of Statistics* **5**, 88–97.

Feynman, R. (2006), *QED: The Strange Theory of Light and Matter*, Princeton, NJ: Princeton University Press.

Field, C., and E. Ronchetti (1990), *Small Sample Asymptotics* (IMS Lecture Notes–Monograph Series No. 13), Hayward, CA: Institute of Mathematical Statistics.

Gbur, E. E., and R. A. Collins (1989), "Estimation of the moment generating function," *Communications in Statistics: Simulation and Computation* **18**, 1113–1134.

Ghosh, J. K., and R. V. Ramamoorthi (2003), *Bayesian Nonparametrics*, New York: Springer-Verlag.

Giffin, W. C. (1975), *Transform Techniques for Probability Modeling*, New York: Academic Press.

Gijbels, I., A. Pope, and M. P. Wand (1999), "Understanding exponential smoothing via kernel regression," *Journal of the Royal Statistical Society: Series B* **61**, 39–50.

Gill, R. D. (1980a), "Nonparametric estimation based on censored observations of a Markov renewal process," *Zeitschrift fur Wahrscheinlickeitstheorie und Verwandte Gebeite* **53**, 97–116.

Gill, R. D. (1980b), *Censoring and Stochastic Integrals* (Mathematical Centre Tracts 124), Amsterdam: Mathematisch Centrum.

Goutis, C., amd G. Casella (1999), "Explaining the saddlepoint approximation," *The American Statistician* **53**, 216–224.

Guihenneuc-Jouyaux, C., S. Richardson, and I. M. Longini (2000), "Modeling markers of disease progression by a hidden Markov process: application to characterizing CD4 cell decline," *Biometrics* **56**, 733–741.

Gusella, V. (1998), "Safety estimation method for structures with cumulative damage," *Journal of Engineering Mechanics* **124**, 1200–1209.

Gusella, V. (2000), "Eigen-analysis of cumulative damage," *Structural Safety* **22**, 189–202.

Hardy, G. H. (1952), *A Course of Pure Mathematics*, 10th Edition, Cambridge: Cambridge University Press.

Hanson, T. E. (2006), "Inference for mixtures of finite Polya tree models," *Journal of the American Statistical Association* **101**, 1548–1565.

Hilbe, J. M. (2006), "Mathematica 5.2: a review," *The American Statistician* **60**, 176–186.

Hjort, N. L. (1996), "Bayesian approaches to non- and semiparametric density estimation," in Bernardo, J. M., *et al.*, *Bayesian Statistics 5: Proceedings of the Fifth Valencia International Meeting, June 5-9, 1994*, Oxford: Oxford University Press, 223–253.

Holton, G. J., and S. G. Brush (2001), *Physics, the Human Adventure: From Copernicus to Einstein and Beyond*, 3rd Edition, Piscataway, NJ: Rutgers University Press.

Hougaard, P. (1999), "Multi-state models: a review," *Lifetime Data Analysis* **5**, 239–264.

Howard, R. A. (1971), *Dynamic Probabilistic Systems Volume I: Markov Models*, New York: John Wiley & Sons (reprinted by Dover Publications, 2007).

Howard, R. A. (1971), *Dynamic Probabilistic Systems Volume II: Semi-Markov and Decision Models*, New York: John Wiley & Sons (reprinted by Dover Publications, 2007).

Huzurbazar, A. V. (2005a), *Flowgraph Models for Multistate Time-to-Event Data*, Hoboken, NJ: Wiley-Interscience.

Huzurbazar, A. V. (2005b), "A censored data histogram," *Communications in Statistics: Simulation and Computation* **34**, 113–120.

Huzurbazar, A.V. (2005c),"Flowgraph models: a Bayesian case study in construction engineering," *Journal of Statistical Planning and Inference* **129**, 181-193.

Huzurbazar, A. V., and B. J. Williams (2005), "Flowgraph models for complex multistate system reliability," in Wilson, A. G., *et al.* (eds.), *Modern Statistical and Mathematical Methods in Reliability*, Singapore: World Scientific.

Huzurbazar, S., and A.V. Huzurbazar (1999), "Survival and hazard functions for progressive diseases using saddlepoint approximations," *Biometrics* **55**, 198–203.

Jensen, J. L. (1995), *Saddlepoint Approximations*, Oxford: Oxford University Press.

Kanai, K. (1983), *Engineering Seismology*, Tokyo: University of Tokyo Press.

Kaplan, E. L., and P. Meier (1958), "Nonparametric estimation from incomplete observations," *Journal of the American Statistical Association*, **53**, 457–481.

Kemeny, J. G., and J. L. Snell (1960), *Finite Markov Chains*, Princeton, NJ: D. Van Nostrand.

Klein, J. P., and M. L. Moeschberger (2003), *Survival Analysis: Techniques for Censored and Truncated Data*, 2nd Edition, New York: Springer-Verlag.

Kleinrock, L. (1975), *Queueing Systems Volume I: Theory*, New York: John Wiley & Sons.

Kolmogorov, A. N. (1956), *Foundations of the Theory of Probability*, New York: Chelsea Publishing (translation of *Grundbegriffe der Wahrscheinlichkeitsrechnung*, 1933).

Kulkarni, V. G. (1996), *Modeling and Analysis of Stochastic Systems*, Boca Raton, FL: Chapman & Hall/CRC.

Künsch, H. R. (2008), "Comment: the 2005 Neyman Lecture: dynamic indeterminism in science," *Statistical Science* **23**, 65–68.

Kvam, P. H., and E. A. Peña (2005), "Estimating load-sharing properties in a dynamic reliability system," *Journal of the American Statistical Association* **100**, 262–272.

Lang, S. (1997), *Undergraduate Analysis*, 2nd Edition, New York: Springer-Verlag.

Law, A. M. and W. D. Kelton (2000), *Simulation Modeling and Analysis*, 3rd Edition, New York: McGraw Hill.

Lee, W. H. K., H. Meyers, and K. Shimazaki, eds. (1988), *Historical Seismograms and Earthquakes of the World*, San Diego: Academic Press.

Limnios, N., and G. Oprişan (2001), *Semi-Markov Processes and Reliability*, Boston: Birkhäuser.

Lindqvist, B. H. (2006), "On the statistical modeling and analysis of repairable systems," *Statistical Science* **21**, 532–551.

Lindsay, B. G. (1995), *Mixture Models: Theory, Geometry, and Applications* (NSF-CBMS Regional Conference Series in Probability and Statistics Volume 5), Hayward, CA: Institute of Mathematical Statistics.

Lorens, C. S. (1964), *Flowgraphs for the Modeling and Analysis of Linear Systems*, New York: McGraw-Hill.

Lukacs, E. (1960), *Characteristic Functions*, London: Charles Griffin & Co.

Lugannani, R. and S. Rice (1980), "Saddle point approximation for the distribution of the sum of independent random variables," *Advances in Applied Probability* **12**, 475–490.

Mallat, S. (1999), *A Wavelet Tour of Signal Processing*, 2nd Edition, San Diego, CA: Academic Press.

Mann, N. R., R. E. Schafer and N. D. Singpurwalla (1974), *Methods for Statistical Analysis of Reliability and Life Data*, New York: John Wiley & Sons.

Mason, S. J. (1953), "Feedback theory—some properties of signal flow graphs," *Proceedings of the I.R.E.* **41**, 1144–1156.

Mason, S. J. (1956), "Feedback theory—further properties of signal flow graphs," *Proceedings of the I.R.E.* **44**, 920–926.

McCabe, T. J. (1976), "A complexity measure," *IEEE Transactions on Software Engineering* **4**, 308–320.

McCormick, N. J. (1981), *Reliability and Risk Analysis: Methods and Nuclear Power Applications*, New York: Academic Press.

Meier, P. (1975), "Estimation of a distribution function from incomplete observations," in Gani, J., ed., *Perspectives in Probability and Statistics: Papers in Honour of M. S. Bartlett*, London: Academic Press, 67–87.

Meyer, D. D. (2000), *Matrix Analysis and Applied Linear Algebra*, Philadelphia, PA: SIAM.

Moore, E. H., and R. Pyke (1968), "Estimation of the transition distributions of a Markov renewal process," *Annals of the Institute of Statistical Mathematics*, **20**, 411–424.

Morcous, G. (2006), "Performance prediction of bridge deck systems using Markov chains," *Journal of Performance of Constructed Facilities* **20**, 146–155.

Neuts, M. F. (1981), *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*, Baltimore: Johns Hopkins University Press (reprinted by Dover Publications, 1995).

Ouhbi, B., and N. Limnios (1999), "Nonparametric estimation for semi-Markov proccesses based on its hazard rate function," *Statistical Inference for Stochastic Processes* **2**, 151–173.

Owen, A. B. (2001), *Empirical Likelihood*, Boca Raton, FL: Chapman & Hall/CRC.

Padgett, W.J. and L.J. Wei (1979), "Estimation for the three-parameter inverse Gaussian distribution," *Communications in Statistics: Theory and Methods* **8**, 129–137.

Pappas, Y. Z., Spanos, P. D., and Kostopoulos, V. (2001), "Markov chains for damage accumulation of organic and ceramic matrix composites," *Journal of Engineering Mechanics* **127**, 915–926.

Parzen, E. (1962), "On estimation of a probability density function and mode," *Annals of Mathematical Statistics* **33**, 1065–1076.

Peña, E. A. (2006), "Dynamic modeling and statistical analysis of event times," *Statistical Science* **21**, 487–500.

Phillips, C. L., and R. D. Harbor (1991), *Feedback Control Systems*, 2nd Edition, Engelwood Cliffs, NJ: Prentice-Hall.

Pyke, R. (1961a), "Markov renewal processes: definitions and preliminary properties," *The Annals of Mathematical Statistics* **32**, 1231–1242.

Pyke, R. (1961b), "Markov renewal processes with finitely many states," *The Annals of Mathematical Statistics* **32**, 1243–1259.

Reid, N. (1988), "Saddlepoint methods and statistical inference," *Statistical Science* **3**, 213-227.

Richter, C. F. (1958), *Elementary Seismology*, San Francisco, CA: W. H. Freeman.

Robichaud, L. P. A., M. Boisvert, and J. Robert (1962), *Signal Flow Graphs and Applications*, Engelwood Cliffs, NJ: Prentice-Hall.

Rosenthal, J. S. (2000), *A First Look at Rigorous Probability Theory*, Singapore: World Scientific.

Ross, S. M. (1970), *Applied Probability Models with Optimization Applications*, San Francisco, CA: Holden-Day (reprinted by Dover Publications, 1992).

Ross, S. M. (1996), *Stochastic Processes*, 2nd Edition, New York: John Wiley & Sons.

Rudin, W. (1976), *Principles of Mathematical Analysis*, 3rd Edition, New York: McGraw-Hill.

Rudin, W. (1987), *Real and Complex Analysis*, 3rd Edition, New York: McGraw-Hill.

Saks, S. (1937), *Theory of the Integral*, G. E. Stechert & Co. (reprinted by Dover Publications, 1964).

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, New York: Wiley-Interscience.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Boca Raton, FL: Chapman & Hall/CRC.

Singh, V.P., J.F. Cruise, and M. Ma (1990), "A comparative evaluation of the estimators of the Weibull distribution by Monte Carlo simulation," *Journal of Statistical Computation and Simulation* **36**, 229–241.

Sprenger, J. (2009), "Statistics between inductive logic and empirical science," *Journal of Applied Logic*, to appear (online doi:10.1016/j.jal.2007.11.007).

Springer, M. D. (1979), *The Algebra of Random Variables*, New York: John Wiley & Sons.

Stigler, S. M. (1986), *The History of Statistics*, Cambridge, MA: Harvard University Press.

Strawderman, R. L. (2004), "Computing tail probabilities by numerical Fourier inversion: the absolutely continuous case," *Statistica Sinica* **14**, 175–201.

Stute, W. (1994), "The bias of Kaplan-Meier integrals," *Scandanavian Journal of Statistics* **21**, 475–484.

Stute, W., and J.-L. Wang (1993), "The strong law under random censorship," *The Annals of Statistics* **21**, 1591–1607.

Takahashi, Y., A. D. Kiureghian, and A. H.-S. Ang (2004), "Lifecycle cost analysis based on a renewal model of earthqhake occurrences," *Earthquake Engineering and Structural Dynamics* **33**, 859–880.

Tapia, R. A., and J. R. Thompson (1978), *Nonparametric Probability Density Estimation*, Baltimore, MD: Johns Hopkins University Press.

Tarter, M. E., and M. D. Lock (1993), *Model-Free Curve Estimation*, New York: Chapman & Hall.

Vere-Jones, D. (1970), "Stochastic models for earthquake occurrence," *Journal of the Royal Statistical Society: Series B* **32**, 1–62.

Wand, M. P., and M. C. Jones (1995), *Kernel Smoothing*, Boca Raton, FL: Chapman & Hall/CRC.

Weideman, J. A. C., (1999), "Algorithms for parameter selection in the Weeks method for inverting the Laplace transform," *SIAM Journal on Scientific Computing* **21**, 111–128.

Widder, D. V. (1946), *The Laplace Transform*, Princeton, NJ: Princeton University Press.

Wiens, D. P., J. Cheng, and N. C. Beaulieu (2003), "A class of method of moments estimators for the two-parameter gamma family," *Pakistan Journal of Statistics*

19, 129–141.

Williams, B. J., and A. V. Huzurbazar (2006), "Posterior sampling with constructed likelihood functions: an application to flowgraph models," *Applied Stochastic Models in Business and Industry* **22**, 127-137.

Wolfram, S. (2003), *The Mathematica Book* (5th ed.), Champaign, IL: Wolfram Media/Cambridge University Press.

Wolstenholme, L. C. (1995), "A nonparametric test of the weakest-link principle," *Technometrics* **37**, 169–175.

Wolstenholme, L. C. (1999), *Reliability Modelling: A Statistical Approach*, Boca Raton, FL: Chapman & Hall/CRC.

Yau, C. L., and Huzurbazar, A. V. (2002), "Analysis of censored and incomplete survival data using flowgraph models," *Statistics in Medicine* **21**, 3727-3743.

Zemanian, A. H. (1987), *Distribution Theory and Transform Analysis: An Introduction to Generalized Functions*, New York: Dover Publications.