Nonparametric Estimation of First Passage Time Distributions in Flowgraph Models

David H. Collins

Department of Mathematics and Statistics
University of New Mexico

Joint work with Aparna V. Huzurbazar
Statistical Sciences Group
Los Alamos National Laboratory

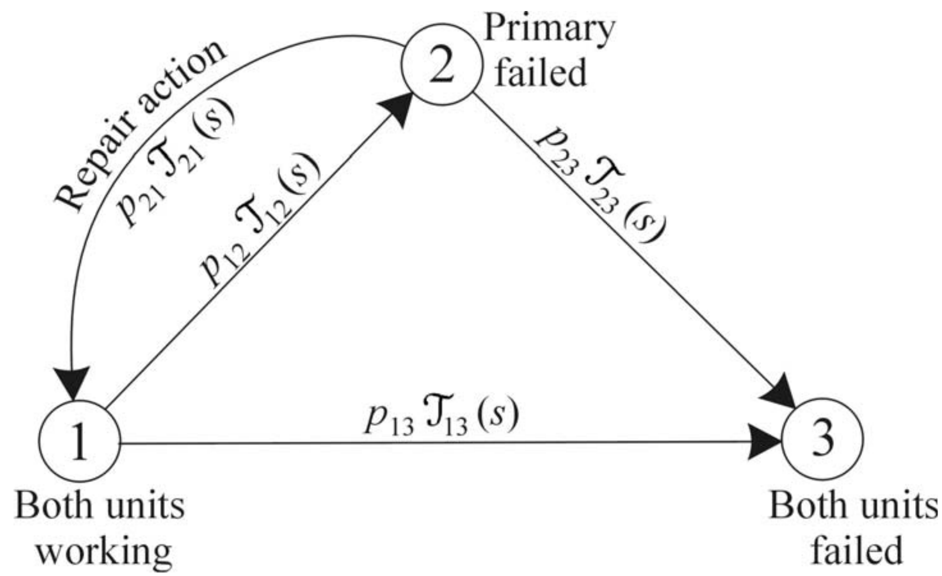
Albuquerque Chapter
American Statistical Association

Outline

- Multistate models
- Overview of parametric statistical flowgraphs
- Nonparametric flowgraph models
  – Empirical transforms
- Reliability applications
  – Repairable redundant system
  - Competing risk of wearout/random failures
  - Detecting a model error
- Pros & cons of nonparametric method
- Summary and conclusions

Multistate model example

- Repairable system with redundancy
  - Or . . . reversible illness-death model
- In general, models of complex systems with many failure modes, feedback (repair) loops

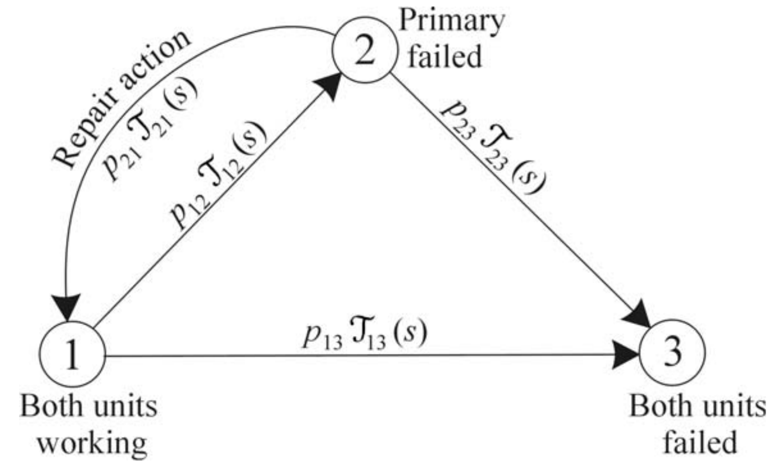


- Interested in time distribution for first passage from "working" to "failing" state – 1 $\rightarrow$ 3 in this example

Formally: time-homogeneous semi-Markov process

- Transition probabilities p_{ij} (embedded Markov chain)
- Arbitrary holding time distributions $F_{ij}(t)$

Overview of "classical" parametric flowgraphs



- <u>Transmittance</u> between states: $p_{ij}\,\mathfrak{T}_{ij}(t)$, $\mathfrak{T}_{ij}(t)$ is MGF or Laplace transform of $F_{ij}(t)$
- Parallel transmittance = sum, series transmittance = product, of branch transmittances
- Transforms reduce solution to an algebraic computation (Mason's rule), e.g., for 1→3 first passage:

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

$$= \frac{p_{01}\,\mathfrak{T}_{01}(s)\,p_{12}\,\mathfrak{T}_{12}(s)}{1 - p_{12}\,\mathfrak{T}_{12}(s)\,p_{21}\,\mathfrak{T}_{21}(s)}$$

- Inversion of computed MGF/LT yields first passage density

A. V. Huzurbazar, *Flowgraph Models for Multistate Time-to-Event Data*, Wiley 2005.

Empirical transforms

<u>Exact transform</u> with kernel ψ of a distribution $F(t)$:

$$\Im(s) = E[\psi(t,s)] = \int \psi(t,s)\,dF(t)$$

Corresponding <u>empirical transform</u> for sample $\{t_1, \ldots t_n\}$ from F:

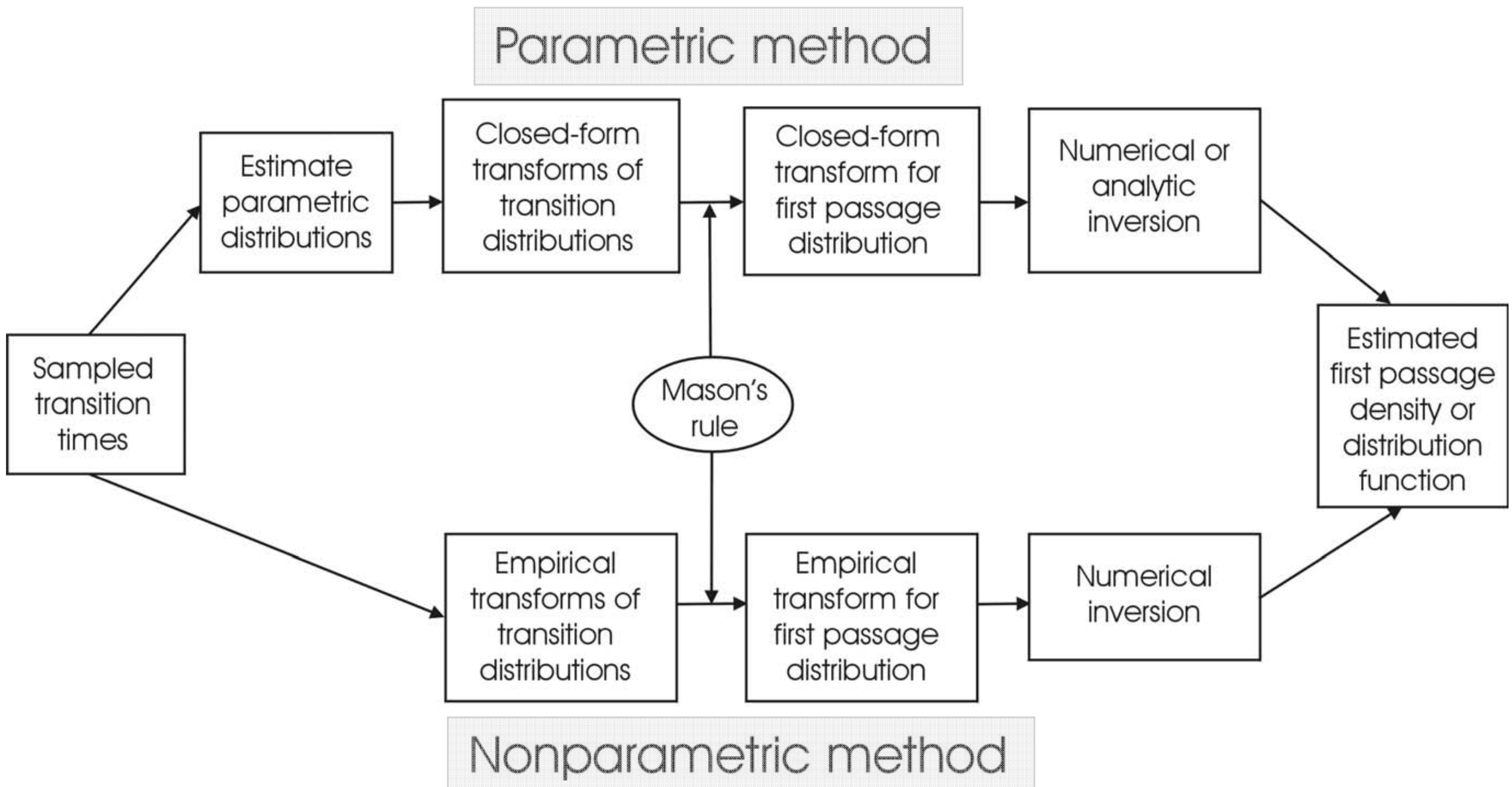
$$\tilde{\Im}(s) = \int \psi(t,s)\,dF_n(t) = \frac{1}{n}\sum_{i=1}^{n}\psi(t_i, s)$$

E.g., Laplace transform (f is the density of F):

$$\mathcal{L}_f(s) = \int_0^\infty e^{-st} f(t;\theta)\,dt, \qquad \tilde{\mathcal{L}}_f(s) = \frac{1}{n}\sum_{i=1}^{n} e^{-st_i}$$

- By SLLN, empirical converges to exact transform
- Retains standard properties, e.g., convolution

Parametric and nonparametric flowgraph methods



Empirical transforms (continued)

- Can mix and match exact and empirical transforms
- Empirical transform for right-censored data can be defined based on Kaplan-Meier estimator
- Inversion methods
  - Saddlepoint approximation based on MGF
  - Fourier series approximation of LT inversion
- Advantages/disadvantages of empirical transforms:
  - Harder to invert (computational complexity)
  - Less accurate then parametric **<u>if</u>** family is known
  - Can be **<u>more</u>** accurate – error is bounded by the data, not the model (see example later)

Numerical inversion of empirical transforms

Objective: recover the density from which a "pseudosample" of first passage times could be drawn

1. Saddlepoint method (Daniels 1954)

$$\hat{f}(t) = \left[\frac{1}{2\pi K''(\hat{s}_t)}\right]^{1/2} \exp\left[K(\hat{s}_t) - \hat{s}_t t\right]$$ where $\hat{s}_t$ (the saddlepoint)

solves $K'(\hat{s}_t) = t$, $\quad K(s) = \log\left[\mathcal{M}(s)\right]$ is the CGF.

Empirical saddlepoint substitutes $\tilde{K}(s) = \log\left[\frac{1}{n}\sum_{i=1}^{n} \exp(st_i)\right]$,

the empirical cgf

Daniels, H. E. (1954), "Saddlepoint approximations in statistics," *The Annals of Mathematical Statistics* 25, 631-650.

Numerical inversion of empirical transforms

2. Approximate the inversion integral with a truncated Fourier series:

$$s_N(t) = \frac{he^{at}}{\pi}\mathrm{Re}\left[\mathcal{L}_f\left(a\right)\right] + \frac{2he^{at}}{\pi}\sum_{k=1}^{N}\mathrm{Re}\left\{\mathcal{L}_f\left(a+ikh\right)\right\}\cos(kht)$$

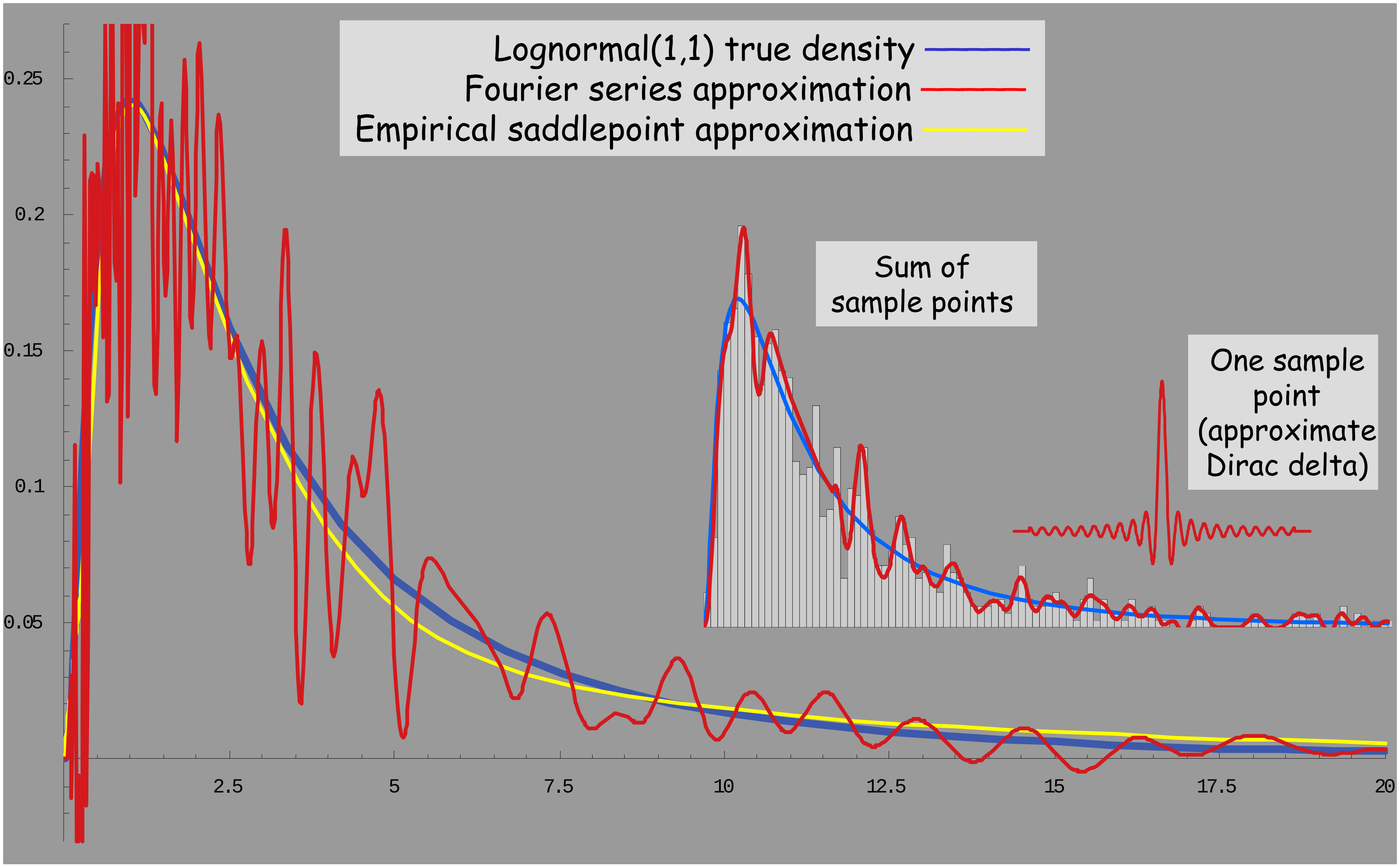
Use Euler summation to accelerate convergence:

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

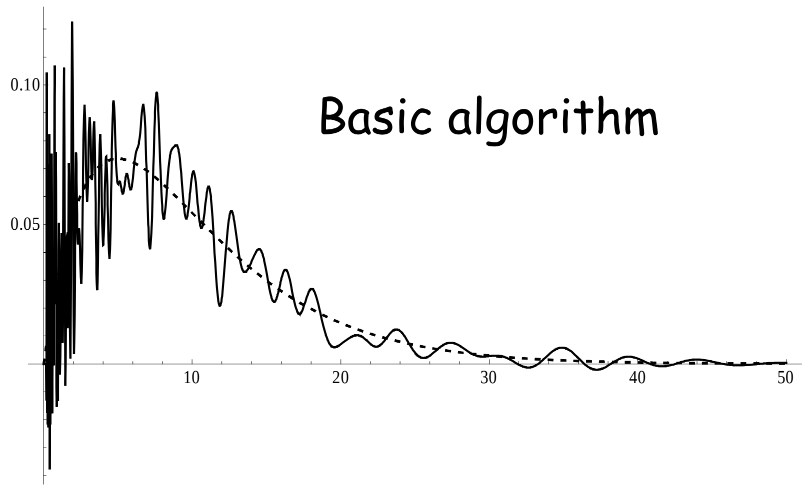
- Very accurate for exact transforms of continuous densities
- Less so for empirical transforms, improves with complexity of flowgraph

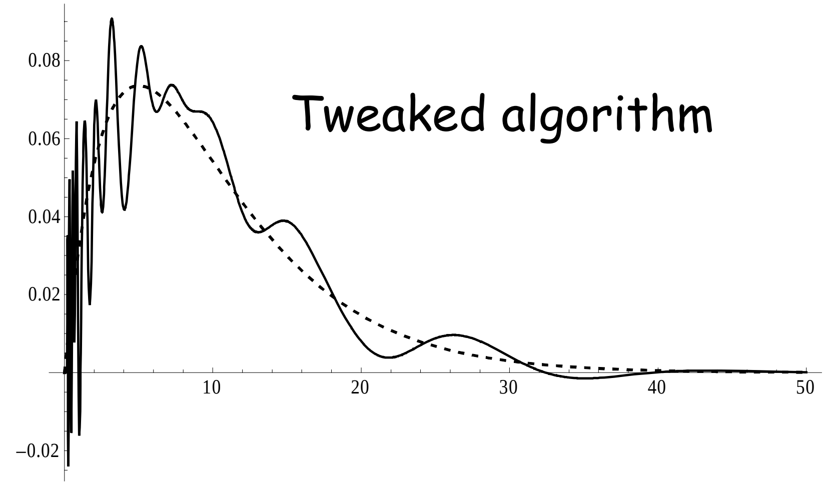
J. Abate and W. Whitt, W. (1992), "The Fourier-series method for inverting transforms of probability distributions," *Queueing Systems* 10, 5-88.

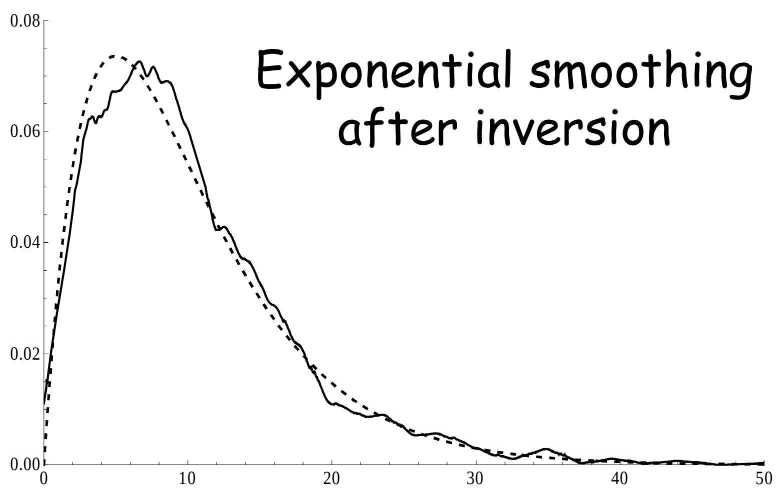
Inversion of an empirical transform

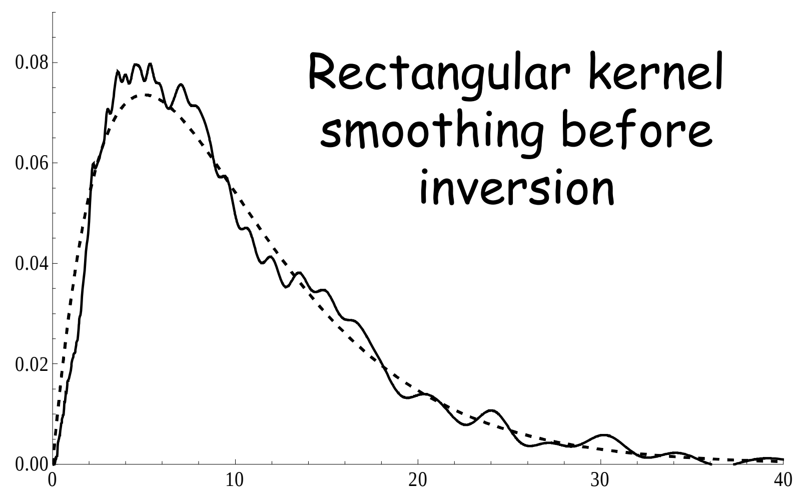


Smoothing the Fourier series approximation







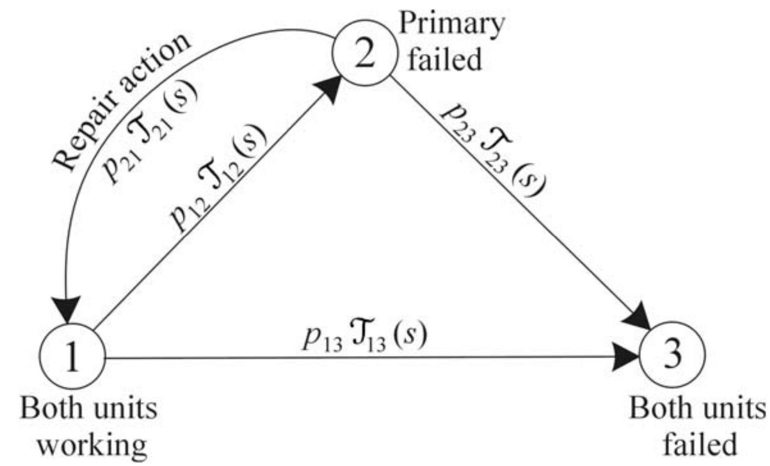


Example – Three-state repairable redundant system

- Each unit has independent failure distribution $\exp(\lambda_1)$ so $1 \to 2$ transition ~ $\exp(2\lambda_1)$ (assume P[simultaneous failure] = 0)

- $1 \to 3$ transition ~ $\exp(\lambda_2)$ (common cause failure, e.g., power outage)

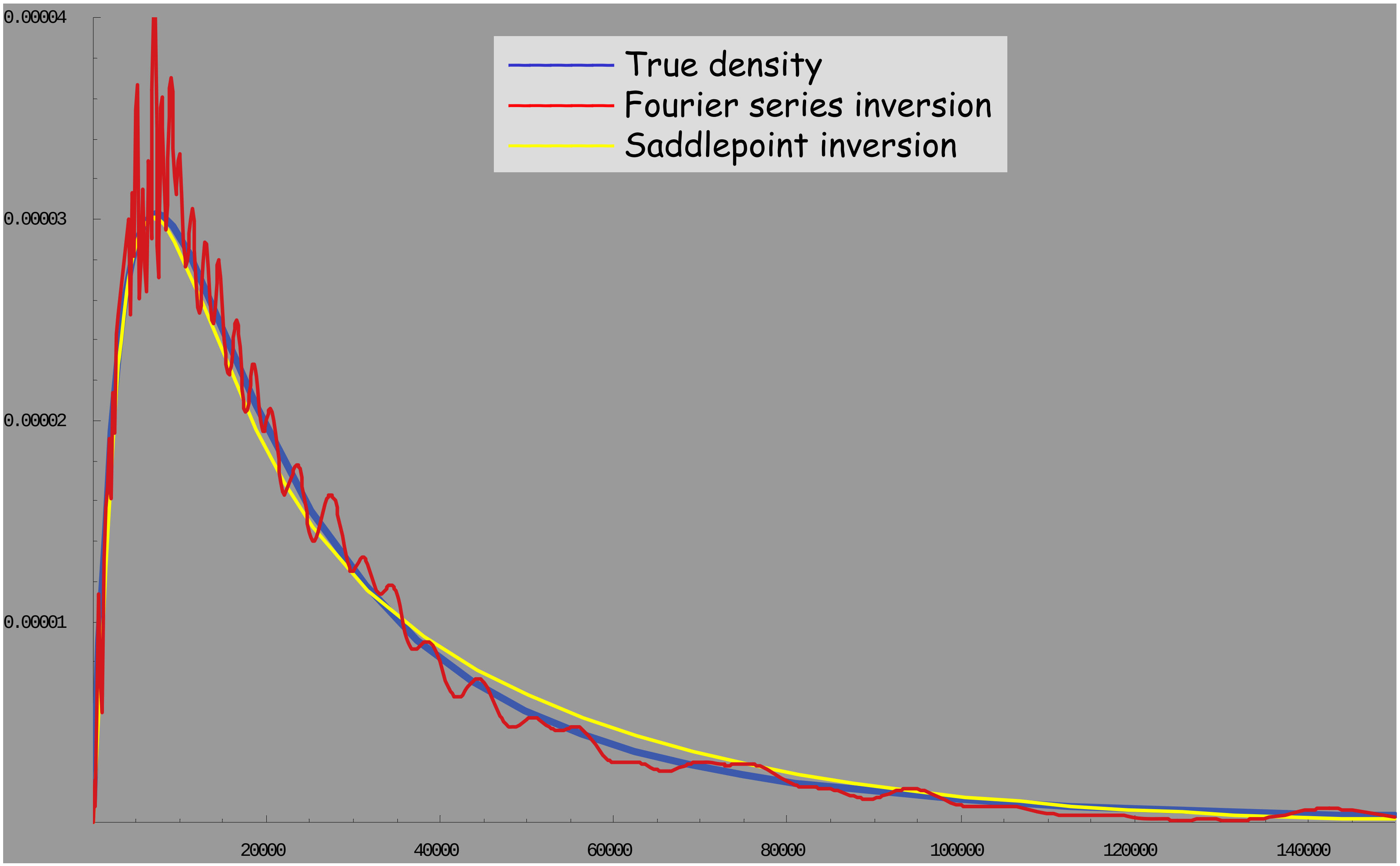
- $2 \to 1$ transition (failed unit repaired) ~ gamma(α, β)



- Sample sizes 270, 25, 247, 23 for $1 \to 2$, $1 \to 3$, $2 \to 1$, $2 \to 3$ (realistic for well-controlled field operation of one system over 1-2 years, or data from multiple systems)

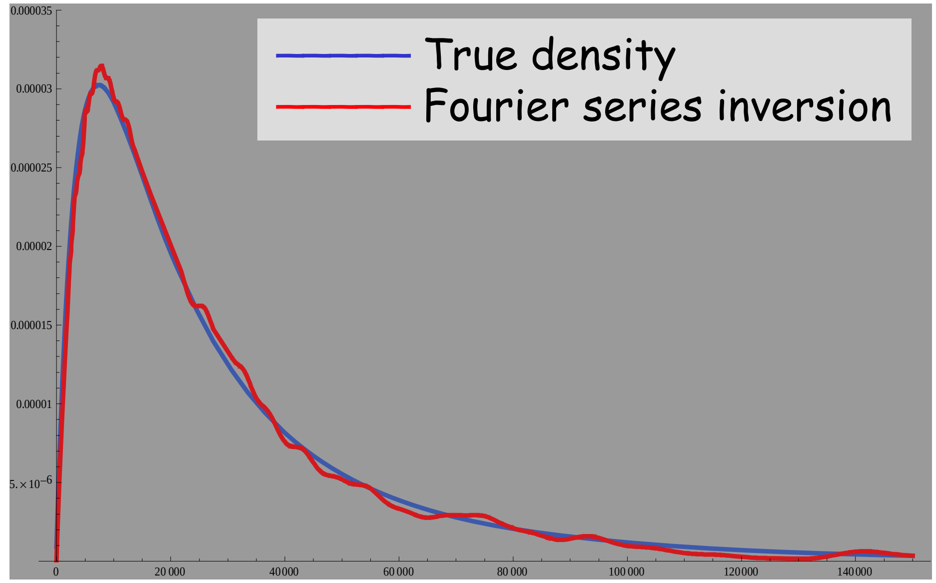
- Interested in $1 \to 3$ first passage $\equiv$ time to total system failure

Three-state repairable redundant system

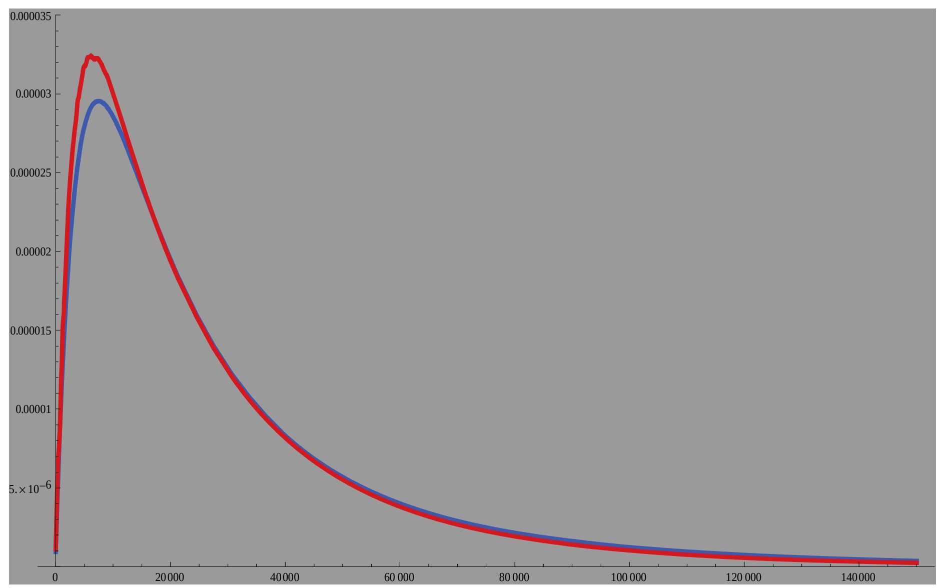


Repairable redundant system - alternatives

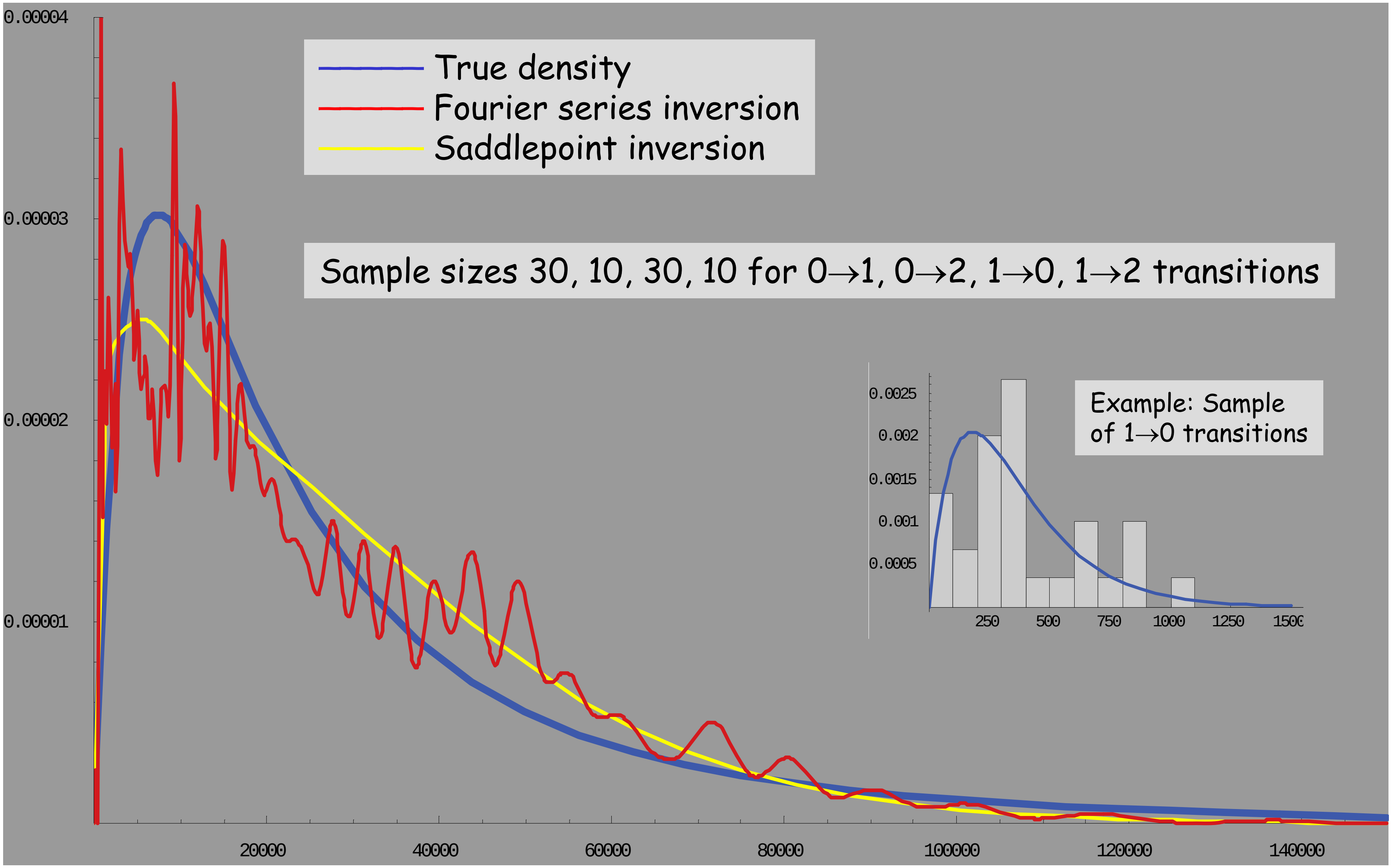
Exponential smoothing of Fourier series inversion points



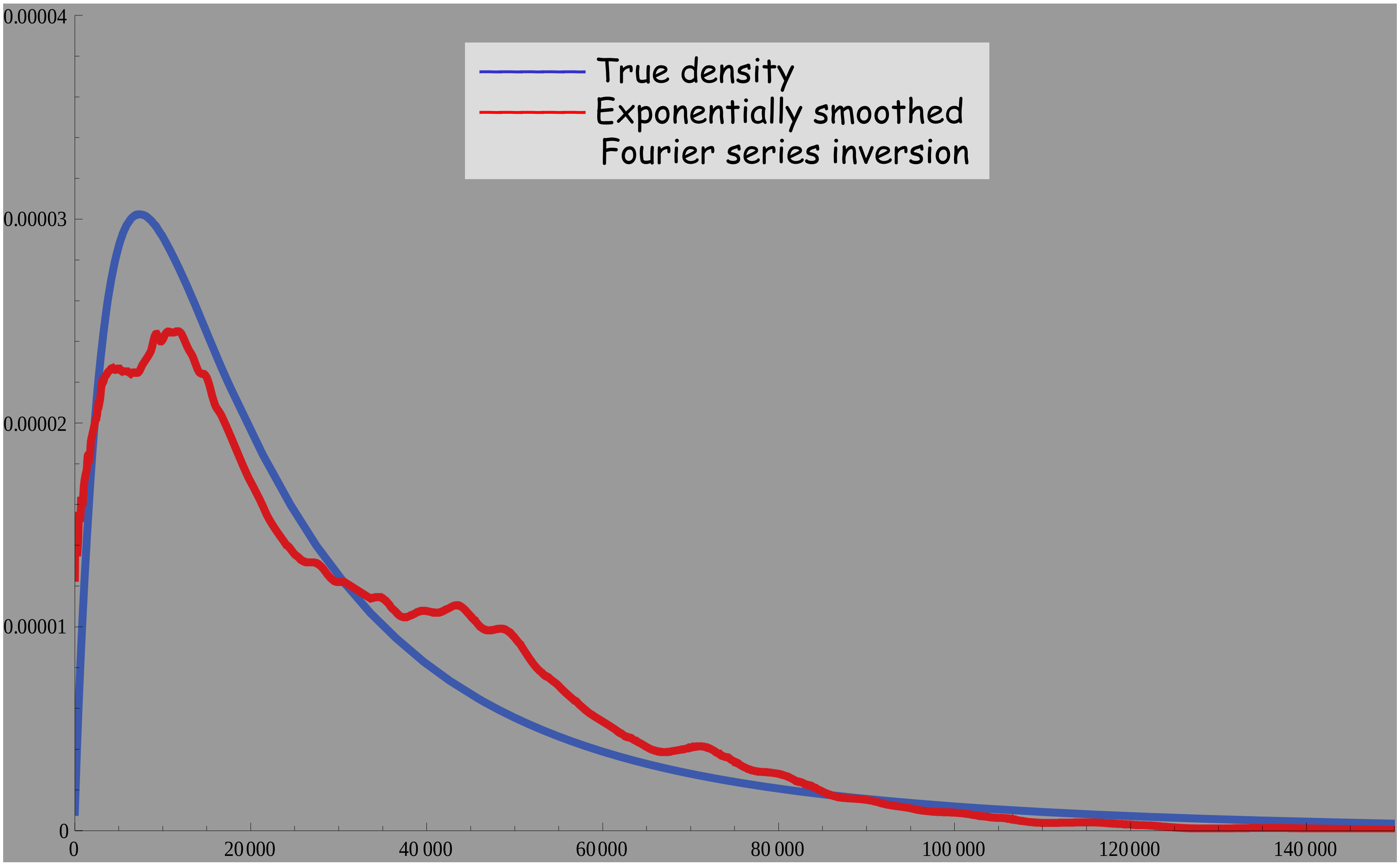
Semiparametric model:
- Empirical transforms for 0→1 and 1→0 transitions
- 0→2 and 1→2 transitions parametric exp(λ), λ estimated by maximum likelihood



Repairable redundant system – small sample

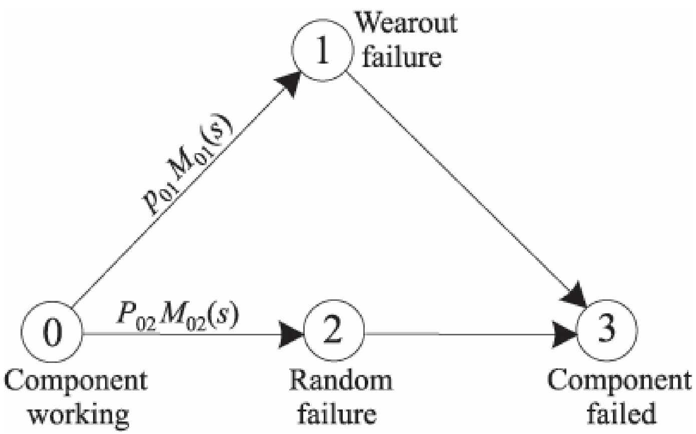


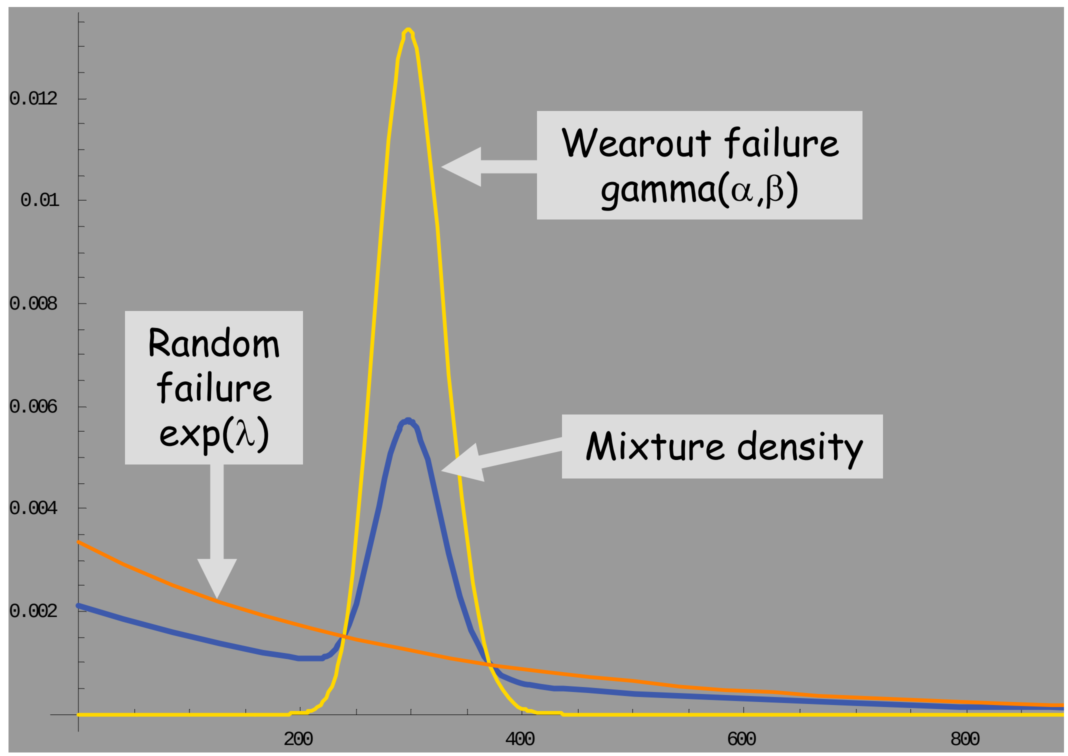
Small sample with exponential smoothing



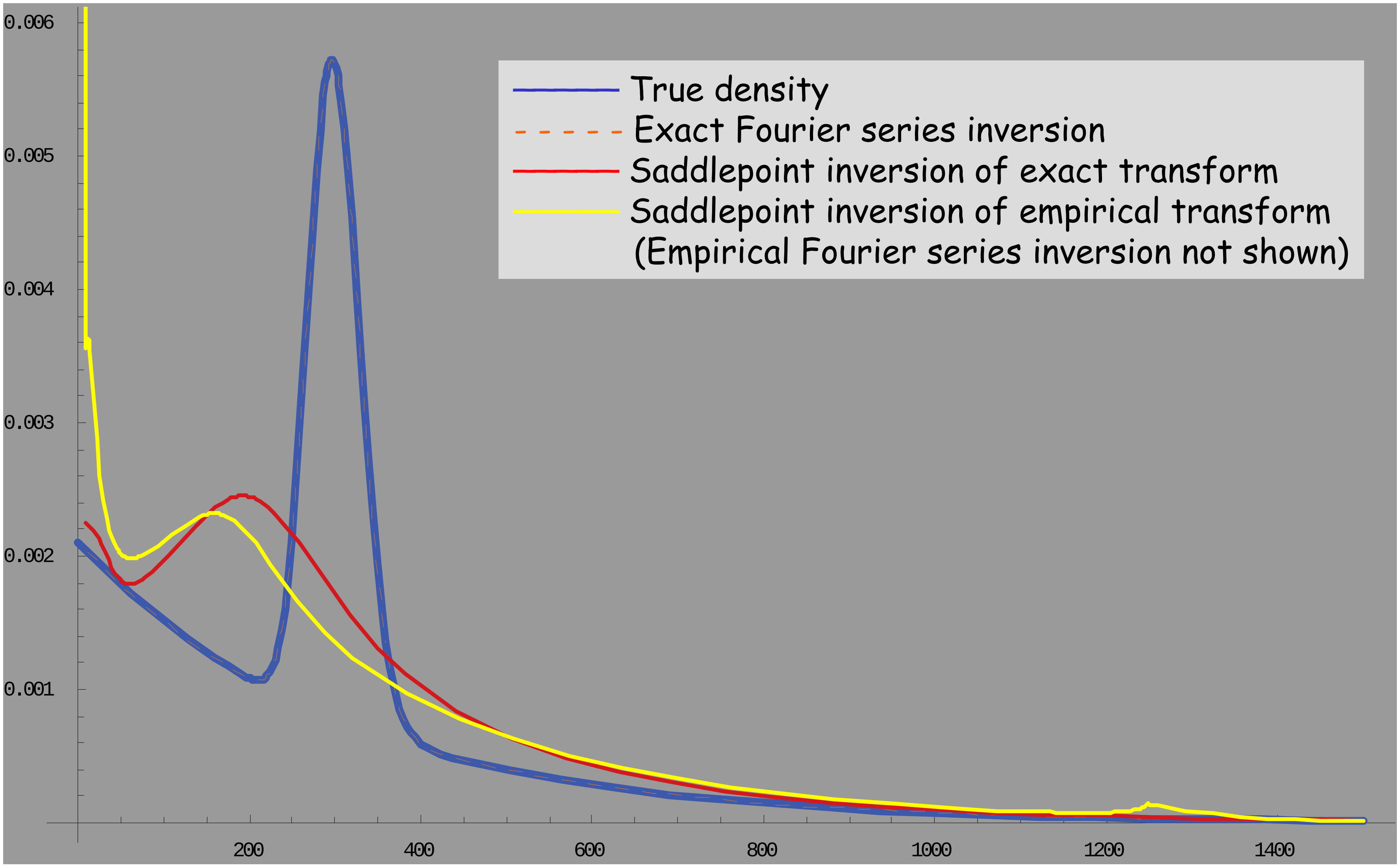
Example – Wearout and random failure modes

- Complex component with two failure modes
- Challenging case for saddlepoint method
- Flowgraph and densities:



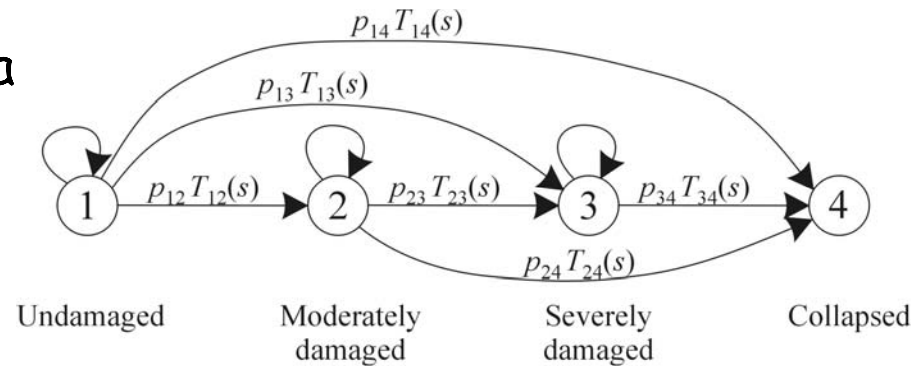


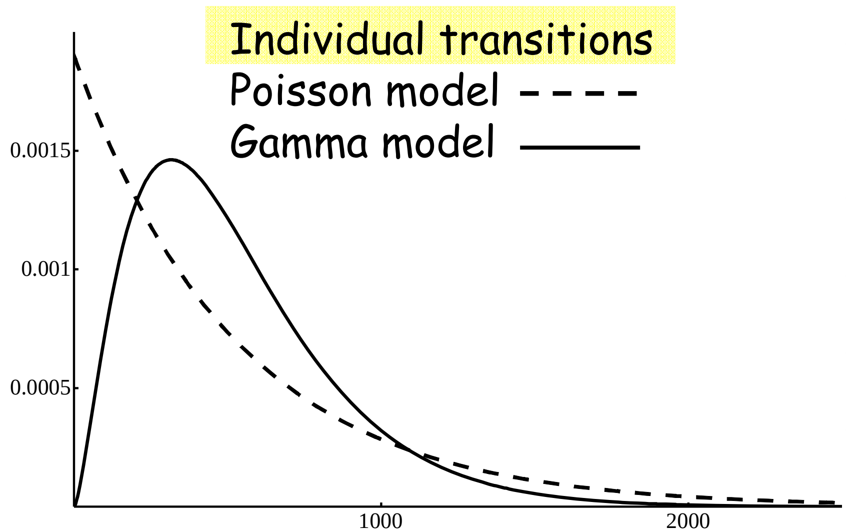
Example – Wearout and random failure modes

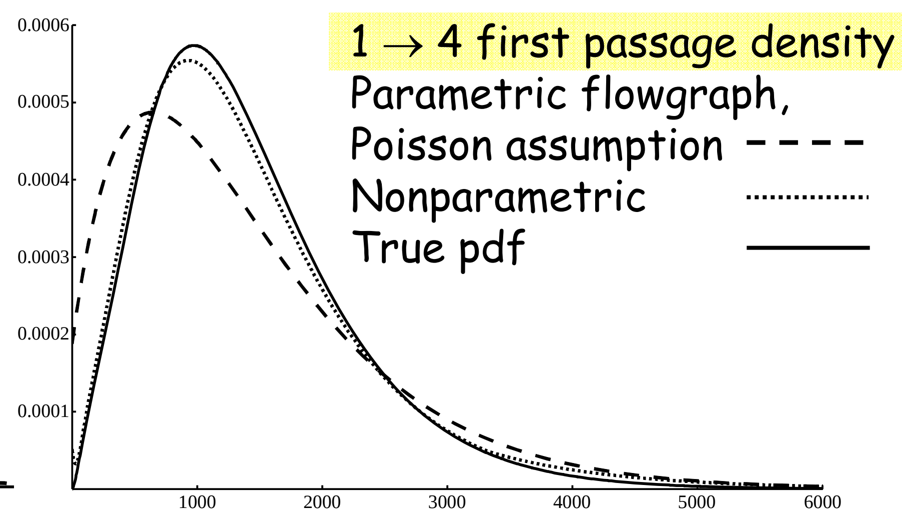


Detecting a model error

Model for cumulative damage to a structure from repeated seismic shocks. Transitions modeled as a homogeneous Poisson process.







D. H. Collins & A. V. Huzurbazar, "System reliability and safety assessment using nonparametric flowgraph models," *Journal of Risk and Reliability*, to appear.

Pros & cons of nonparametric method

- How much confidence do we have in a particular parametric model?
  - Non-parametric result can be worse, or better
  - Small-sample results may not be very useful
- Computational complexity is much greater with non-parametric method (probably improvable)
- Empirical transform always exists (e.g., for lognormal)
- Non-parametric method is a valuable adjunct even if a parametric model is available
  - Reinforcement if confirmatory
  - If not, indicates need for further analysis

Summary and conclusions

- Illustrated nonparametric method for statistical flowgraphs
  - Realistic examples for reliability scenarios
  - Useful in itself, and for validation of parametric results
- Further work:
  - More efficient algorithms (e.g., for numerical integration of empirical saddlepoint)
  - Adaptive smoothing for Abate & Whitt inversion
  - Handling of left/interval censored/truncated data
  - Large-sample properties of nonparametric approximations

Thanks for your attention . . .
Questions?