

Estimation of System Reliability Using a Semiparametric Model

Leon Wu, *Member, IEEE*, Timothy Teräväinen, Gail Kaiser, *Senior Member, IEEE*,
Roger Anderson, *Member, IEEE*, Albert Boulanger, Cynthia Rudin

Abstract—An important problem in reliability engineering is to predict the failure rate, that is, the frequency with which an engineered system or component fails. This paper presents a new method of estimating failure rate using a semiparametric model with Gaussian process smoothing. The method is able to provide accurate estimation based on historical data and it does not make strong *a priori* assumptions of failure rate pattern (e.g., constant or monotonic). Our experiments of applying this method in power system failure data compared with other models show its efficacy and accuracy. This method can be used in estimating reliability for many other systems, such as software systems or components.

Index Terms—estimation theory, failure analysis, Gaussian processes, parametric statistics, power system reliability, prediction methods, reliability engineering, software reliability, statistical analysis, stochastic processes.

I. INTRODUCTION

RELIABILITY is one of the most important requirements of the smart grid and other sustainable energy systems. By *smart grid*, we refer to an automated electric power system that monitors and controls grid activities, ensuring the two-way flow of electricity and information between power plants and consumers—and all points in between [1]. In the past ten years, the U.S. power grid has become less reliable and more failure-prone; according to two data sets, one from the U.S. Department of Energy and the other one from the North American Electric Reliability Corp., the number of power outages greater than 100 Megawatts or affecting more than 50,000 customers in the U.S. almost doubles every five years, resulting in about \$49 billion outage costs per year [2].

How to accurately and effectively evaluate system reliability has been a long-time research challenge. One commonly used indicator for system reliability is *failure rate*, which is the frequency with which an engineered system or component fails. To estimate the failure rate, historical failure information and/or testing of a current sample of equipment are commonly used as the basis of the estimation. After these data have

Leon Wu is with the Department of Computer Science and the Center for Computational Learning Systems, Columbia University, New York, NY 10027 USA (e-mail: leon@ccls.columbia.edu).

Timothy Teräväinen is with the Department of Statistics, Columbia University, New York, NY 10027 USA (e-mail: tkt2103@columbia.edu).

Gail Kaiser is with the Department of Computer Science, Columbia University, New York, NY 10027 USA (e-mail: kaiser@cs.columbia.edu).

Roger Anderson and Albert Boulanger are with the Center for Computational Learning Systems, Columbia University, New York, NY 10027 USA (e-mail: {anderson, aboulanger}@ccls.columbia.edu).

Cynthia Rudin is with the MIT Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02142 USA (e-mail: rudin@mit.edu).

been collected, a failure distribution model, *i.e.*, a cumulative distribution function that describes the probability of failure up to and including time t , is assumed (e.g., the exponential failure distribution or more generally, the Weibull distribution) and used to estimate the failure rate.

Our experimental results indicate that using an exponential or Weibull distribution prior may not be as effective for power grid failure modeling as a particular semiparametric model introduced in this work. This semiparametric model does not assume a constant or monotonic failure rate pattern as the other models do. We introduce Gaussian smoothing that further helps the semiparametric model to closely resemble the true failure rate. We applied this method to power network component failure data and compared its blind-test estimation results with the subsequent real failures. We also compared it with other models during these experiments. In all of these cases, the semiparametric model outperformed the other models.

The paper is organized as follows. In the following section, we will present some background information on reliability analysis. Then we will describe our new model in detail, followed by experimental results and analysis. We will further compare our approach with other models. We will conclude the paper after discussing related work.

II. BACKGROUND ON RELIABILITY ANALYSIS

The failure rate can be defined as the total number of failures within an item population, divided by the total time expended by that population, during a particular measurement interval under stated conditions [3]. We use $\lambda(t)$ to denote the failure rate at time t , and $R(t)$ to denote the reliability function (or survival function), which is the probability of no failure before time t . Then the failure rate is:

$$\lambda(t) = \frac{R(t) - R(t + \Delta t)}{\Delta t \cdot R(t)}.$$

As Δt tends to zero, the above λ becomes the instantaneous failure rate, which is also called hazard function (or hazard rate) $h(t)$:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{R(t) - R(t + \Delta t)}{\Delta t \cdot R(t)}.$$

A failure distribution $F(t)$ is a cumulative failure distribution function that describes the probability of failure up to and including time t :

$$F(t) = 1 - R(t), t \geq 0.$$

For system with a continuous failure rate, $F(t)$ is the integral of the failure density function $f(t)$:

$$F(t) = \int_0^t f(x) dx.$$

Then the hazard function becomes

$$h(t) = \frac{f(t)}{R(t)}.$$

A. Weibull and Exponential Failure Distribution

For the Weibull failure distribution, the failure density function $f(t)$ and cumulative failure distribution function $F(t)$ are

$$f(t; \lambda, k) = \begin{cases} \frac{k}{\lambda} \left(\frac{t}{\lambda}\right)^{k-1} e^{-(t/\lambda)^k}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

$$F(t; \lambda, k) = \begin{cases} 1 - e^{-(t/\lambda)^k}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

where $k > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter of the distribution. The hazard function when $t \geq 0$ can be derived as

$$h(t; \lambda, k) = \frac{f(t; \lambda, k)}{R(t; \lambda, k)} = \frac{f(t; \lambda, k)}{1 - F(t; \lambda, k)} = \frac{k}{\lambda} \left(\frac{t}{\lambda}\right)^{k-1}.$$

A value of $k < 1$ indicates that the failure rate decreases over time. A value of $k = 1$ indicates that the failure rate is constant (*i.e.*, k/λ) over time. In this case, the Weibull distribution becomes an exponential distribution. A value of $k > 1$ indicates that the failure rate increases with time.

III. SEMIPARAMETRIC MODEL WITH GAUSSIAN SMOOTHING

We consider the semiparametric estimation of the longitudinal effect of a blip treatment (*i.e.*, a single “all-or-nothing” treatment occurring at a precisely recorded time) on a system with recurring events (*e.g.*, immediately-recoverable failures in a mechanical/electronic system). The estimand is the effect of the most recent blip treatment on the future arrival rate. The method assumes that the effect of treatment is to scale the underlying rate, and is thus an extension of Cox regression with internal covariates, using the Gaussian process to provide much-needed smoothing.

Although the method applies to any blip treatment, we focus on estimating the effect of an event (failure) on future failures. For example, an association of an event with an immediate increase in failure rate provides a finely-detailed explanation for “infant mortality” which can be compared with parametric models such as the Weibull.

A. Probability and Regression Model

We assume each of N units is under observation for some interval of time $[0, T]$. The method can be easily adapted to allow for units with missing observation periods (known in advance). Let \mathbb{T} denote the (finite) set of times at which an event occurs. The unit to fail at time t (if any) is denoted as $i(t)$; ties are broken in preprocessing, if necessary, by randomly selecting tied units and shifting their failures by one

second. For any unit j under observation at time t denote by $\tau_{t,i}$ the time of the treatment (which is here the time of previous outage). It turns out to be important to remove “unobserved” units (*i.e.* those for which $t - \tau_{t,i}$ is unknown due to left-truncation of the study); thus, the index-set of fully-observed units at time t is given by $\mathfrak{R}(t)$, and commonly called the “risk set.” Note that if the mechanism for observation is independent of the treatment and failure processes (*i.e.*, if it is fixed in advance), this does not introduce bias [4]. We consider the non-parametric rate model as follows:

$$\lambda(t; i) = \lambda_0(t) \psi(t - \tau_{t,i});$$

$$\psi(\cdot) = e^{\phi(\cdot)},$$

that is, 20 seconds after treatment the effect will be to make failure $\psi(20) = e^{\phi(20)}$ times more likely.

The full likelihood is then [4]:

$$l(\lambda_0(\cdot), \psi(\cdot)) = \left(\prod_{t \in \mathbb{T}} \lambda_0(t) \psi(t - \tau_{t,i(t)}) \right) \times e^{-\int_0^T \sum_{j \in \mathfrak{R}(t)} \lambda_0(t) \psi(t - \tau_{t,j}) dt}.$$

The estimation proceeds in two steps, detailed in Appendix B. The λ_0 term is first shown to be estimated as 0 at all times $t \notin \mathbb{T}$. Thus, conditioning on the failure times, the λ_0 term is cancelled out (since it affects all units equally). This allows convenient estimation of $\psi(t) = e^{\phi(t)}$. After the estimation of $\psi(t)$, the λ_0 term may be estimated by a weighted non-parametric estimator (which uses the estimate of ψ). For simplicity, in this paper we fit the λ_0 as a constant (within each network) by using the method of moments (Appendix C).

Since only the time since last treatment is tracked, it is implicitly assumed that any prior treatments are immediately “forgotten” by the system upon administration of a new treatment.

The connection between the hazard λ and the distribution function is detailed in Appendix A.

The information reduction induced by the Cox framework should be very useful, especially in the Gaussian process setup which scales as $O(p^3)$ in the number of predictors. To achieve further reduction of data for numerical stability and to expedite cross-validation, we “bin” values of $t - \tau_{t,\cdot}$ (which can be viewed as the predictors of $\phi(t - \tau_{t,\cdot})$) into percentiles.

B. Application

The method is applied to the failure rate of distribution power feeders in three boroughs of New York City (Manhattan, Queens, and Brooklyn). Distribution feeders are the power cables that feed intermediate voltage power in distribution grids. In New York City, underground distribution feeders, mostly 27KV or 13KV, are one of the most failure-prone electrical components in the power grid. The effect of infant mortality and the changing hazard rate are of interest for maintenance scheduling applications.

In our application, $N = 81$ and there are $|\mathbb{T}| = T = 667$ distinct failure times (*i.e.*, 667 total failures are observed among the 81 units).

C. Preliminary Fit

The model predictions without smoothing are provided in Figure 1, which shows the failure rate versus time since treatment, and they are clearly overfitted to the data. Since events occur rarely, we have that some $(t - \tau_{t,i})$ -bins may be observed only once, associated with a failure, causing a direct estimate of $\psi(\cdot)$ to overestimate. Likewise, many bins will be associated only with the non-failed risk set, and $\psi(\cdot)$ will go to 0. This effect will be more pronounced with a large number of units and rare failures.

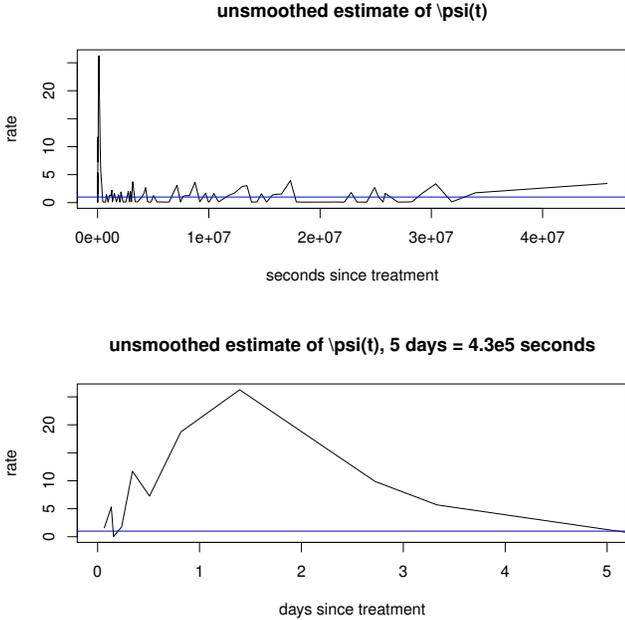


Fig. 1: Preliminary fit

D. Gaussian Process

We apply a Gaussian process prior to the values $\phi(t)$ with a radial basis function. After the standard marginalizing of the prior [5] onto $t \in \mathbb{T}$, the $\phi(t)$ are normally distributed with mean 0 and covariance matrix K with

$$K_{t,t'} = ae^{-(t-t')^2/b}.$$

This marginal prior distribution will be referred to as π . The parameters a, b are the marginal variance and so-called “characteristic time-scale” respectively. We use the parameter values $a = 5$, $b = 1 \cdot 10^3$ based on good performance on the training data. Alternatively, cross-validation on a grid search on these parameters can be used to obtain approximate “point estimates” of a, b .

Details of the fitting process are in Appendix D.

Figure 2 shows the smoothed fit using the Gaussian process prior. It is much better than the unsmoothed fit.

IV. EMPIRICAL STUDY

We implemented the semiparametric model with Gaussian smoothing and applied it to five years of distribution power feeder failure data collected in New York City, as discussed in Section III B. We further compared the estimation with what actually happened. We also applied the exponential distribution

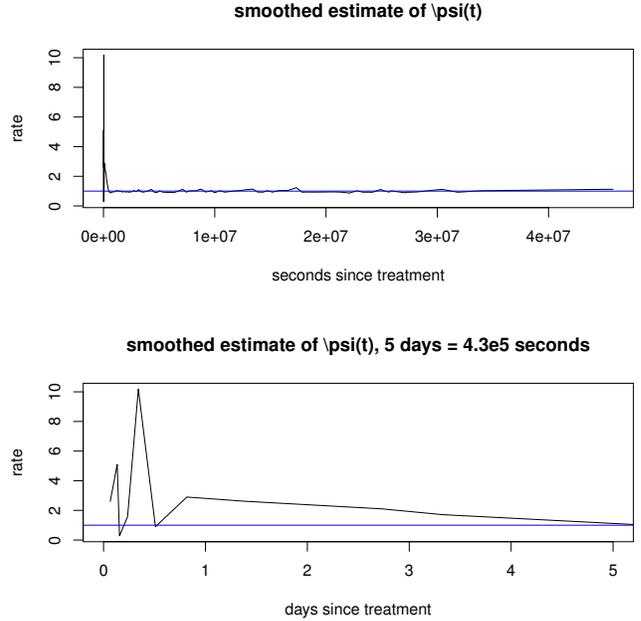


Fig. 2: Smoothed fit

and Weibull distribution models on the same set of data and compared their results with the results from the semiparametric model.

A. Experimental Setup

Our experiments consist of three main groups of blind tests. In New York City, the distribution power feeder failures are seasonal. During summer heat waves, more feeder failures are likely to happen. The three groups are estimates of the failure rate for the summer, winter, and the whole year using the historical data for the first three years, *i.e.*, from year 2006 through 2008. Then we compare these estimates with the actual failure rates measured for the years 2009–2010 using the failure data. We perform similar experiments on the exponential and Weibull models.

B. Results and Analysis

The results of fitting the model are summarized in Table I (giving the constants) and Figure 3 (giving the estimated failure rate multiplier $\psi(t)$) for each network.

TABLE I: Summary of results (units are in days)

Network	# of Units	# of Failures	Exponential λ
Queens: 01Q	26	327	75.2
Brooklyn: 01B	29	197	154.12
Manhattan: 02M	26	143	114.1
Network	Weibull k	Weibull λ	Semiparametric λ_0
Queens: 01Q	0.48	42	71.0
Brooklyn: 01B	0.69	120.4	130.0
Manhattan: 02M	0.62	108.0	112.1

To analyze the fit of each model, we integrate (numerically for the semiparametric) to convert the hazard estimates to estimates of the cumulative distribution function (see Section 3 and Appendix A). The resulting model fits are then visually and numerically compared to the empirical distribution function of the data.

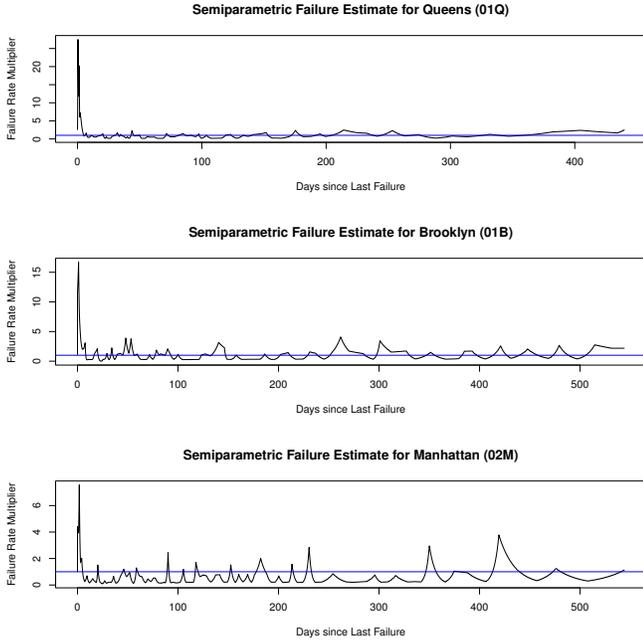


Fig. 3: Semiparametric infant mortality rate estimates

TABLE II: Kolmogorov-Smirnoff test of fit

Training			
Network	Exponential	Weibull	Semiparametric
Queens: 01Q	0.40	0.19	0.13
Brooklyn: 01B	0.25	0.17	0.14
Manhattan: 02M	0.27	0.17	0.12
Testing			
Network	Exponential	Weibull	Semiparametric
Queens: 01Q	0.35	0.23	0.20
Brooklyn: 01B	0.27	0.20	0.16
Manhattan: 02M	0.38	0.31	0.32

The fit of each model is evaluated on the training (2006–08) and test (2009–10) sets using the Kolmogorov-Smirnoff (K-S) statistic [6], which is a distance between the empirical distribution of the cumulative distribution function F , \hat{F}_{emp} , and the F provided by each model fit. Since none of these models are to be considered true, we use the statistic simply as a “measure of fit” on training and holdout data, rather than as a formal hypothesis test. The empirical distribution function is defined as

$$\hat{F}_{\text{emp}}(t) = \frac{1}{T} \sum \mathbb{I}_{t_i < t},$$

with the sum being over all inter-arrival times in the data. The K-S statistic is the maximum absolute discrepancy between the two distributions, defined as

$$\text{KS}(\hat{F}_{\text{emp}}, F) = \sup_t |\hat{F}_{\text{emp}}(t) - F_{\text{model}}(t)|.$$

As expected, the Weibull uniformly performs better than the exponential. Table II shows the K-S test of fit. The semiparametric method uniformly outperforms the Weibull on the training data, and outperforms the Weibull on the holdout test data in Queens and Brooklyn, demonstrating accuracy in prediction. The semiparametric method comes very close to the Weibull in test performance on the Manhattan network which, notably, also exhibits the worst degradation from training to test performance, across all models.

The comparison of the estimation results shows that the failure rate estimates using the semiparametric model are closer to the actual measured inter-arrival times, which means the semiparametric model with Gaussian smoothing is more accurate in estimating the failure rate.

V. RELATED WORK

Estimation of system reliability by modeling failure rate has been an active research area. Various estimation models have been proposed for different kinds of systems including power electrical components, semiconductor chips and boards, and software systems. The exponential, Bayesian, log-normal, and Weibull approaches were popular in prior research. In 1974, Littlewood and Verrall used a Bayesian reliability model to estimate stochastic monotone failure rates [3]. Ibrahim et al. formalized the field of Bayesian survival analysis in 2001 [7]. Rigdon and Basu described a way to estimate the intensity function of a Weibull process [8]. Mudholkar and Srivastava used the exponential Weibull family for analyzing the bathtub failure rate model [9]. In prior sections, we compared our approach with the exponential and Weibull models. Our approach differs from previous Bayesian models in making fewer assumptions on a continuous failure distribution.

Among the failure patterns, the bathtub model and infant mortality are perhaps the most well-studied [10], [9]. To model non-constant failure rates, Jones used a constant failure intensity assumption and exponential failure distribution-based method to do the estimation, and experimented with the method in reliability analysis of digital circuit devices [11]. Our approach does not assume a constant failure rate or a constant failure intensity. The semiparametric model we described is not a modified version of the exponential or Weibull models.

In 1984, Laprie described a mathematical model for the failure behavior of component-based software systems with physical and design faults [12]. Hierons and Wiper researched the estimation of software system failure rate using random and partition testing methods [13]. Kubal et al. proposed a way of estimating software system failure rate based on the failure rates of the underlying components using a Bayesian approach [14]. Although we directly applied our approach to distribution power feeder failures here, our approach can be directly applied to other areas, for instance, software reliability analysis.

VI. CONCLUSION

This paper presented a new method of estimating failure rate using a semiparametric model with Gaussian process smoothing. The method is able to provide accurate estimation based on historical data and it does not make strong *a priori* assumptions of the failure rate pattern (e.g., constant or monotonic). Our empirical studies of applying such an approach in power system failure data and a comparison of this approach with other existing models show its efficacy and accuracy. This method may also be used in estimating reliability for many other systems, such as software systems or components.

APPENDIX A
EQUIVALENCE OF HAZARD AND DISTRIBUTION
FUNCTIONS

From definition of the hazard function,

$$\lambda(t) = f(t)/(1 - F(t)),$$

and from the definitions of

$$f(t) = -\frac{\partial(1 - F(t))}{\partial t},$$

and finally, from calculus,

$$\frac{\partial \log(f(t))}{\partial t} = \frac{f'(t)}{f(t)}.$$

Therefore:

$$\begin{aligned} \frac{-\frac{\partial(1-F(t))}{\partial t}}{1-F(t)} &= \lambda(t), \\ -\frac{\partial \log(1-F(t))}{\partial t} &= \lambda(t), \\ -\log(1-F(t)) &= \int_0^t \lambda(u) du, \\ 1-F(t) &= e^{-\int_0^t \lambda(u) du}, \\ F(t) &= 1 - e^{-\int_0^t \lambda(u) du}. \end{aligned}$$

APPENDIX B
MARGINALIZING TIMES WITHOUT FAILURE

We consider the contribution to the likelihood from the observation of no failures between times t_{i-1}, t_i , assuming no censoring and that $\phi(\cdot) < \infty$:

$$L = e^{-\int_{t_{i-1}}^{t_i} \lambda_0(u) \sum_{j \in \mathfrak{R}(u)} e^{\phi(i-\tau_{u,j})} du}.$$

Taking the functional derivative of λ_0 at time $s \in (t_{i-1}, t_i)$:

$$\frac{\partial L}{\partial \lambda_0(s)} = \left(e^{-\int_{t_{i-1}}^{t_i} \lambda_0(u) \sum_{j \in \mathfrak{R}(u)} e^{\phi(i-\tau_{u,j})} du} \right) \times \left(-\lambda_0(s) \sum_j e^{\phi(s-\tau_{s,j})} \right),$$

which is negative for all positive values of $\lambda_0(s)$. Since $\lambda_0 \geq 0$ by definition, the maximum likelihood estimate of baseline hazard is $\hat{\lambda}_0(s) = 0$, which gives the MLE (*i.e.*, Maximum Likelihood Estimation) of failure rate

$$\hat{\lambda}_0(s) \sum_j e^{\phi(s-\tau_{s,j})} = 0.$$

Substituting this into the likelihood, we see that it does not depend on ϕ when there are no failures, reducing the estimation problem to event times. This result, derived more formally [15], is also valid under random censoring, as shown by Cox and given in [4].

Thus, since intervals without failures give no information about ϕ , we can reduce the problem of estimating ϕ to the conditional probability of each *observed unit failing at time t*, given that *some unit failed at time t*, which is:

$$\prod_t \frac{\text{unit } i \text{ fails at } t}{\text{some unit fails at } t}$$

$$\begin{aligned} &= \prod_t \frac{\lambda_0(t) e^{\phi(t-\tau_{t,i})}}{\lambda_0(t) \sum_j e^{\phi(t-\tau_{t,j})}} \\ &= \prod_t \frac{e^{\phi(t-\tau_{t,i})}}{\sum_j e^{\phi(t-\tau_{t,j})}}, \end{aligned}$$

which gives the ‘‘Cox likelihood’’ for ϕ at those values $t - \tau_{t,j}$, which are observed.

After the estimate of ϕ is obtained, we can derive an estimate of $\Lambda_0 = \int_0^t \lambda_0$ through the weighted non-parametric Nelson-Aalen estimator [16]. This Λ_0 is smoothed and used directly in computing the test-penalty, or if desired λ_0 may be approximately estimated by differentiating the smoothed version.

APPENDIX C
FITTING λ_0

For simplicity we take the baseline hazard λ_0 to be constant for each network. After estimating ψ , the reliability function is

$$R(t) = e^{-\int_0^t h(u)} = e^{-\lambda_0 \int_0^t \psi(u) du},$$

from which the mean time to failure can be computed directly by the so-called layered representation of the expectation (which follows from integration by parts):

$$\mathbb{E}_{\lambda_0}[T] = \int_0^\infty e^{-\lambda_0 \int_0^u \psi(u) du} du dt.$$

At this point, the λ_0 is chosen by grid search over numeric approximations of this integral, so that the mean time to failure equals the empirical mean time to failure: $\mathbb{E}_{\lambda_0}[T] = \bar{T}$.

APPENDIX D
FITTING THE GAUSSIAN PROCESS

The log-posterior probability is proportional to the sum of the log of the Cox likelihood (l) and the log of the marginalized Gaussian process prior (π):

$$\begin{aligned} \frac{\partial L}{\partial \lambda_0(s)} = l + \pi = & \sum_t \log \phi(t - \tau_{t,i}) - \\ & \log \sum_{j \in \mathfrak{R}(t)} \phi(t - \tau_{t,j}) + \\ & \left(-\frac{1}{2} \phi^\dagger K^{-1} \phi \right). \end{aligned}$$

We apply the Newton-Raphson method to find the maximum a-posteriori estimate. The gradient with respect to ϕ is

$$\nabla(l + \pi) = \sum_t \frac{-\psi(t - \tau_{t,i}) + e_{i(t)} s_t}{s_t} + K^{-1} \phi,$$

with Hessian

$$(\nabla \nabla(l + \pi))_{i,j} = \psi(t - \tau_{i,t}) \psi(t - \tau_{j,t}) / s_t^2 + K^{-1},$$

where

$$s_t = \sum_j \psi(t - \tau_{j,t}),$$

the total hazard of observed units at time t , and $e_{i(t)}$ is the unit basis vector indicating the failed unit at time t , $\delta_{i(t)}$.

The step-size is dynamically adjusted, and is stopped on a relative improvement of the quasi-posterior probability by less than $1.4e - 08$.

REFERENCES

- [1] Federal Smart Grid Task Force, 2010, *Smart grid basics*, available: <http://www.smartgrid.gov/basics>.
- [2] S. M. Amin, "U.S. electrical grid gets less reliable," *IEEE Spectrum*, vol. 48, no. 1, pp. 80, Jan 2011.
- [3] B. Littlewood and J. L. Verrall, "A Bayesian Reliability Model with a Stochastically Monotone Failure Rate," *IEEE Transactions on Reliability*, vol. R-23, no. 2, pp. 108-114, 1974.
- [4] P. K. Andersen, O. Borgan, R. D. Gill, N. Keiding, *Statistical Models Based on Counting Processes*, Springer, 1995.
- [5] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, 2006.
- [6] R.H.C. Lopes, I. Reid, and P.R. Hobson, "The two-dimensional Kolmogorov-Smirnov test," in *XI International Workshop on Advanced Computing and Analysis Techniques in Physics Research*, Amsterdam, April 2007.
- [7] J. G. Ibrahim, M.-H. Chen, and D. Sinha, *Bayesian Survival Analysis*, Springer, 2001.
- [8] S. E. Rigdon and A. P. Basu, "Estimating the intensity function of a Weibull process at the current time: Failure truncated case," *Journal of Statistical Computation and Simulation (JSCS)*, vol. 30, pp. 17-38, 1988.
- [9] G. S. Mudholkar and D. K. Srivastava, "Exponentiated Weibull family for analyzing bathtub failure-rate data," *IEEE Transactions on Reliability*, vol. 42, no. 2, pp. 299-302, 1993.
- [10] J. Mi, "Bathtub failure rate and upside-down bathtub mean residual life," *IEEE Transactions on Reliability*, vol. 44, no. 3, pp. 388-391, 1995.
- [11] J. Jones and J. Hayes, "Estimation of system reliability using a "non-constant failure rate" model," *IEEE Transactions on Reliability*, vol.50, no.3, pp. 286-288, Sep 2001.
- [12] J.-C. Laprie, "Dependability evaluation of software systems in operation," *IEEE Transactions on Software Engineering*, vol. SE10, no. 6, pp. 701714, 1984.
- [13] R. M. Hierons and M. P. Wiper, "Estimation of Failure Rate using Random and Partition Testing," *Software Testing, Verification & Reliability*, vol. 7, no. 3, pp. 153-164, 1997.
- [14] S. Kubal, J. May, and G. Hughes, "Building a system failure rate estimator by identifying component failure rates," in *Proc. of International Symposium on Software Reliability Engineering (ISSRE)*, 1999.
- [15] T. Martinussen and T. H. Scheike, *Dynamic Regression Models for Survival Data*, Springer, 2006.
- [16] J. Kalbfleisch and R. Prentice, *The Statistical Analysis of Failure Time Data*, Wiley-Interscience, 2002.

Leon Wu (M'07) is a PhD Candidate at the Department of Computer Science and a senior research associate at the Center for Computational Learning Systems of Columbia University. He received his MS and MPhil in Computer Science from Columbia University and BSc in Physics from Sun Yat-sen University.

Timothy Teräväinen is a PhD student at the Department of Statistics of Columbia University. He received his MPhil in Statistics from Columbia University and his B.A. in Mathematics from New College of Florida.

Gail Kaiser (M'85-SM'90) is a Professor of Computer Science and the Director of the Programming Systems Laboratory in the Computer Science Department at Columbia University.

She was named an NSF Presidential Young Investigator in Software Engineering and Software Systems in 1988, and she has published over 150 refereed papers in a range of software areas. Her research interests include software testing, collaborative work, computer and network security, parallel and distributed systems, self-managing systems, Web technologies, information management, and software development environments and tools. She has consulted or worked summers for courseware authoring, software process and networking startups, several defense contractors, the Software Engineering Institute, Bell Labs, IBM, Siemens, Sun and Telcordia. Her lab has been funded by NSF, NIH, DARPA, ONR, NASA, NYS Science & Technology Foundation, and numerous companies. Prof. Kaiser served on the editorial board of IEEE Internet Computing for many years, was a founding associate editor of ACM Transactions on Software Engineering and Methodology, chaired an ACM SIGSOFT Symposium on Foundations of Software Engineering, vice chaired three of the IEEE International Conference on Distributed Computing Systems, and serves frequently on conference program committees. She also served on the Committee of Examiners for the Educational Testing Service's Computer Science Advanced Test (the GRE CS test) for three years, and has chaired her department's doctoral program since 1997. Prof. Kaiser received her PhD and MS from CMU and her ScB from MIT.

Roger Anderson (M'09) has been at Columbia University for 35 years, where he is a Senior Scholar at the Center for Computational Learning Systems in the Fu School of Engineering and Applied Sciences (SEAS). Roger is Principal Investigator of a team of 15 scientists and graduate students in Computer Sciences at Columbia who are jointly developing the next generation Smart Grid for intelligent control of the electric grid of New York City with Con Edison and others in New York City. Previously at the Lamont-Doherty Earth Observatory of Columbia, Roger founded the Borehole Research, Global Basins, 4D Seismic, Reservoir Simulation, Portfolio Management, and Energy Research Groups. Roger also teaches Planet Earth, a science requirement course in the core curriculum at Columbia College from his position in the Department of Earth and Environmental Sciences. He co-founded the Alternative Energy program at the School of International and Public Affairs at Columbia, and is a director of the Urban Utility Center at the Polytechnic Institute of New York University.

Roger received his Ph.D. from the Scripps Institution of Oceanography, University of California at San Diego. He is the inventor of 16 Patents, and has written 3 books, & more than 200 peer-reviewed scientific papers. In addition to his desk at the Manhattan Electric Control Center of Con Edison for the last 7 years, he has had technical, business, computational, and working collaborations with many other companies, including Baker Hughes, Boeing, BBN, BP, Chevron, IBM Research, KBR, Lockheed Martin, Pennzoil, Schlumberger, Siemens, Shell, United Technologies, and Western GECO. Rogers specialties include the Smart Grid, Optimization of Control Center Operations of Energy Companies, Real Options and Portfolio Management, 4D Reservoir Management, and Alternative Energy Research. His new book on the subject, Computer-Aided Lean Management, from PennWell Press, is available on Amazon.com. He has written scientific and opinion pieces for magazines such as CIO Insight, Discover, Economist, EnergyBiz, Forbes, National Geographic, Nature, New York Times, Oil and Gas Journal, Scientific American, Wall Street Journal, and Wired. Roger assisted in the design of the Wiess Energy Hall at the Houston Museum of Natural History, was technical consultant for the NBC News/Discovery Channel documentary Anatomy of a Blackout, and has been a frequent contributor to business radio and TV.

Albert Boulanger received a B.S. in physics at the University of Florida, Gainesville, Florida USA in 1979 and a M.S. in computer science at the University of Illinois, Urbana-Champaign, Illinois USA in 1984.

He is a co-founder of CALM Energy, Inc. and a member of the board at the not-for-profit environmental and social organization World Team Now and founding member of World-Team Building, LLC. He is a Senior Staff Associate at Columbia University's Center for Computational Learning Systems, and before that, at the Lamont-Doherty Earth Observatory. For the past 12 years at Columbia, Albert has been involved in far reaching energy research and development in oil and gas and electricity He is currently a member of a team of 15 scientists and graduate students in Computer Sciences at Columbia who are jointly developing with Con Edison and others the next generation Smart Grid for intelligent control of the electric grid of New York City. He held the CTO position of vPatch Technologies, Inc., a startup company commercializing a computational approach to efficient production of oil from reservoirs based on time-lapse 4D seismic technologies. Prior to coming to Lamont, Albert spent twelve years doing contract R&D at Bolt, Beranek, and Newman (now Raytheon BBN Technologies). His specialties are complex systems integration and intelligent computational reasoning that interacts with humans within large scale systems.

Cynthia Rudin is an Assistant Professor of Statistics at the MIT Sloan School of Management, and an Adjunct Research Scientist at the Center for Computational Learning Systems at Columbia University. Her research interests are in statistical learning theory and applications of machine learning. She has several projects that involve the application of machine learning to electrical grid maintenance, and also a project with Ford through the MIT-Ford Alliance. She is the recipient of an NSF CAREER award and an NSF Postdoctoral Research Fellowship. She received a PhD at Princeton University, and BS and BA degrees from the State University of New York at Buffalo.