Journal of Multivariate Analysis 101 (2010) 340–351

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Wiener processes with random effects for degradation data

Xiao Wang

Department of Mathematics and Statistics, University of Maryland, 1000 Hilltop Circle, Baltimore, MD 21250, USA

ARTICLE INFO

Article history:
Available online 24 December 2008

AMS subject classifications:
62G20
62M05
62N05

Keywords:
Bootstrap
Degradation
EM algorithm
Empirical processes
Random effects
Reliability
Wiener process

ABSTRACT

This article studies the maximum likelihood inference on a class of Wiener processes with random effects for degradation data. Degradation data are special case of functional data with monotone trend. The setting for degradation data is one on which n independent subjects, each with a Wiener process with random drift and diffusion parameters, are observed at possible different times. Unit-to-unit variability is incorporated into the model by these random effects. EM algorithm is used to obtain the maximum likelihood estimators of the unknown parameters. Asymptotic properties such as consistency and convergence rate are established. Bootstrap method is used for assessing the uncertainties of the estimators. Simulations are used to validate the method. The model is fitted to bridge beam data and corresponding goodness-of-fit tests are carried out. Failure time distributions in terms of degradation level passages are calculated and illustrated.

1. Introduction

In some studies where the subjects are put on test at time zero, these subjects degrade over time. Usually continuous observation of degradation for these subjects is not possible. The degradation of each subject may be observed at each of several times, where the number of observation times and observation times themselves are allowed to vary across the subjects. Data of this type are called *degradation data*. Degradation data are special case of functional data with monotone trend. Degradation data have applications in many fields, such as HIV study and industrial reliability. For example, it is suggested that the immune system of a person infected with the HIV virus degrades over time [1]. CD4 counts constitute a critical assessment of the status of the immune system and CD4 counts are commonly used markers for the health status of HIV infected persons. In reliability area, the data from fatigue crack growth subject to loading cycles [2,3] and bridge beams subject to erosion of chloride ion ingression [4] are typical degradation data. Degradation data are a very rich source of survival information. In many tests, the failure time data are supplemented by degradation data and degradation data offer many advantages over failure time data. For general discussion of degradation models, see [5–7].

Wiener processes and extensions to them have been used as models for degradation data (e.g. [8–11]). Let $\Lambda(t)$ be a nondecreasing function. The nonhomogeneous Wiener process $Y(t)$ has independent increments $\Delta Y(t) = Y(t+\Delta t) - Y(t)$, where $\Delta Y(t)$ has a normal distribution with mean $\Delta\Lambda(t) = \Lambda(t+\Delta t) - \Lambda(t)$ and variance $\sigma^2\Delta\Lambda(t)$. If letting $U(t) = t + \sigma W(t)$ be the Wiener process with drift t and diffusion σ, then

$$Y(t) = U(\Lambda(t)) = \Lambda(t) + \sigma W(\Lambda(t)) \tag{1}$$

is just a time-transformed Wiener process. When the amount of degradation reaches a pre-specified critical level D, failure occurs. Let T denote the failure time, then $T = \inf\{t : Y(t) \geq D\}$. The level-crossing of the cumulative degradation threshold D by a nonhomogeneous Wiener process $Y(t)$ can be obtained in terms of the inverse Gaussian (IG) distribution [12,13,31] as

E-mail address: wangxiao@umbc.edu.

0047-259X/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2008.12.007

$T_\Lambda = \Lambda^{-1}(T_{IG})$, where T_{IG} is inverse Gaussian (IG) random variable $IG(D, D^2/\sigma^2)$ with mean D and variance σ^2/D. Note the similarities between nonhomogeneous Wiener process model and hazard rate model. The failure time in hazard rate model can be obtained by $T_\Lambda = \Lambda^{-1}(T_{\exp})$ where Λ is the integrated hazard rate and $T_{\exp}$ is the standard exponential distribution. Detailed nonparametric inference of this model is discussed by [14].

In most degradation applications, there is substantial subject-to-subject variability among the degradation processes of different individuals. The unit-specific random effects can be incorporated into the process to represent such heterogeneity in the degradation paths. One such model can be specified by allowing both drift and diffusion of the above Wiener process model to be random. Mathematically tractable distribution results if we adopt a model as follows:

$$\text{Given } (\nu, \sigma), \quad Y(t) = \nu\Lambda(t) + \sigma W(\Lambda(t)), \tag{2}$$

$$w = \sigma^{-2} \sim \mathrm{Gam}(r^{-1}, \delta), \qquad \nu | w \sim N(1, \theta/w). \tag{3}$$

Here, w has mean δ/r and variance δ/r^2 and thus σ^2 has finite mean $r/(\delta-1)$ for $\delta > 1$ and finite variance $r^2/((\delta-1)^2(\delta-2))$ for $\delta > 2$. Letting the conditional mean of ν be 1 is for model identification since any constant can be incorporated into function $\Lambda(t)$. An alternative parameterization that is often used in such situations is to let $r = \delta$ in the distribution of w. There is no physical meaning for choosing the random effect distributions in (3). The idea of making these choices of the distributions (3) for random effects is borrowed from Bayesian linear regression [15] for computational convenience. Semiparametric maximum likelihood method is developed to estimate the unknown parameter $(\Lambda(t), \delta, r, \theta)$ of the underlying process. Here, $\Lambda(t)$ is estimated nonparametrically, which leads naturally to an infinite dimensional statistical problem. We show that the maximum likelihood estimator (MLE) is consistent and we also derive the convergence rate of the MLE. Bootstrap method is used for assessing the uncertainty of the MLE. Simulation results suggest that this method works well and we apply our method to the degradation data of a civil engineering structure to estimate its reliability.

The remainder of the paper is as follows. Section 2 introduces the Wiener process model with random effects. Section 3 establishes the consistency and convergence rate of the maximum likelihood estimator and uses bootstrap method to assess the uncertainties of the estimators. Section 4 uses EM algorithm to compute the MLE. Section 5 presents Monte Carlo studies to validate the methods and we also fit the Wiener process model to bridge beam data in Section 6. Section 7 makes some concluding remarks.

2. Wiener process with random effects: *t* process

Model (2) shows that the conditional distribution of $Y(t)$ given w and ν is normal, and the marginal density of $Y(t)$ follows as

$$\begin{aligned} f(y) &= \int_{-\infty}^{\infty}\int_0^{\infty} f(y; \nu, w) g_1(\nu; \theta, w) g_2(w; r, \delta) \mathrm{d}w \mathrm{d}\nu \\ &= \frac{\Gamma(\delta + \frac{1}{2})}{\sqrt{2\pi r}\,\Gamma(\delta)[\Lambda^2\theta + \Lambda]^{1/2}} \left[1 + \frac{(y - \Lambda)^2}{2r(\Lambda^2\theta + \Lambda)}\right]^{-\delta - \frac{1}{2}}. \end{aligned} \tag{4}$$

Note that $\sqrt{\frac{\delta}{r(\Lambda^2\theta+\Lambda)}}(Y(t) - \Lambda(t))$ has a t distribution with degrees of freedom 2δ. Thus, $Y(t)$ has finite mean $\Lambda(t)$ and finite variance

$$\mathrm{Var}(Y(t)) = [\Lambda(t)^2\theta + \Lambda(t)]\frac{r}{\delta - 1}, \quad \text{for } \delta > 1.$$

An extreme situation of the above model is when the variances of the random effects are zero and it becomes the general Wiener process model (1). This situation can be realized by letting $\theta \to 0$ and $r \to \infty$ with $\delta/r = c$ fixed. Conditionally on their common random effects, ν and w, the level-crossing of the cumulative degradation threshold D by the process $Y(t)$ in (2) follows IG distribution $IG(D/\nu, D^2 w)$. Hence, $P(T \le t) = E_{\nu,w} G(\Lambda(t); \nu, w)$, where $G(t, \nu, w)$ is the inverse Gaussian distribution function with parameters $(D/\nu, D^2 w)$. If the degradation paths are monotonic, the failure time distribution has explicit form and is given by

$$P(T \le t) = P(Y(t) > d) = F_{2\delta}\left[\sqrt{\frac{\delta}{r}}\frac{\Lambda(t) - d}{\sqrt{\theta\Lambda(t)^2 + \Lambda(t)}}\right], \tag{5}$$

where $F_{2\delta}$ is the t distribution function with degrees of freedom 2δ.

Suppose that we observe $Y(t)$ for a subject at times $t_1, \ldots, t_m$, yielding observations $Y_1, \ldots, Y_m$. Let $y_j = Y_j - Y_{j-1}$ and $\lambda_j = \Lambda(t_j) - \Lambda(t_{j-1})$ with $Y_0 = 0$ and $\Lambda(t_0) = 0$. Given the common random effects ν and w, the increments y_j are independently normally distributed. Thus the joint density of the y_j can be obtained as

$$\begin{aligned} f(y_1, \ldots, y_m) &= \int_{-\infty}^{\infty}\int_0^{\infty} \phi(y_1, \ldots, y_m; \nu, w) g_1(\nu; \theta, w) g_2(w; r, \delta) \mathrm{d}w \mathrm{d}\nu \\ &= \frac{\Gamma(\delta + \frac{m}{2})}{\Gamma(\delta)(\sqrt{2\pi r})^{m/2}|A|^{1/2}}\left[1 + \frac{1}{2r}(y - \lambda)' A^{-1}(y - \lambda)\right]^{-\delta - \frac{m}{2}}, \end{aligned} \tag{6}$$

where $y = (y_1, \dots, y_m)'$ and $\lambda = (\lambda_1, \dots, \lambda_m)'$ and the elements of matrix A are given by $[A]_{ii} = \lambda_i^2\theta + \lambda_i$ and $[A]_{ij} = \lambda_i\lambda_j\theta$ for $i \neq j$. The joint density (6) for $(y_1, \dots, y_m)'$ is a multivariate noncentral t distribution with degrees of freedom 2δ and covariance matrix $rA/(\delta - 1)$, for $\delta > 1$, with $\mathrm{cov}(y_j, y_k) = \theta r\lambda_j\lambda_k/(\delta - 1)$. We call such process $Y(t)$ a *t process*. One of the important issues to study degradation data is to predict the residual failure time conditional on observed levels of degradation. As pointed out by [16], if we know the current degradation measurement, $P(T > t+\Delta t|T > t, Y(t))$ provides more precise prediction than $P(T > t+\Delta t|T > t)$. To track the individual subject, we need estimates of the random effects. One such estimates can be obtained by

$$E[\nu|Y_1, \dots, Y_m] = \frac{\theta^{-1} + \sum_{j=1}^{m} y_j}{\theta^{-1} + \sum_{j=1}^{m} \lambda_j}, \tag{7}$$

and

$$E[w|Y_1, \dots, Y_m] = \left(\delta + \frac{m}{2}\right)\left[-\frac{(\theta^{-1} + \sum_{j=1}^{m} y_j)^2}{2(\theta^{-1} + \sum_{j=1}^{m} \lambda_j)} + \sum_{j=1}^{m} \frac{y_j^2}{2\lambda_j} + \frac{1}{2\theta} + r\right]^{-1}. \tag{8}$$

Given the random effects ν and w, the residual failure time distribution, conditional on the last observed level of degradation being $Y(t_k)$ at t_k, is just the first passage time to $D - Y(t_k)$ of the nonhomogeneous Wiener process with drift function $\Lambda^*(t) = \Lambda(t) - \Lambda(t_k)$.

Suppose that we observe the degradation paths for subject i at some discrete times t_{ij}, i.e., we observe $Y_{i,j} = Y_i(t_{ij}), j = 1, \dots, m_i, i = 1, \dots, n$. This type of data can be treated as a special case of functional data with monotone trend. The simplest situation is that the observation times are fixed and the same for all subjects, say $t_1, \dots, t_m$. Under this circumstance, we just observe a sample of multivariate t random variables $(Y_{i,1}, \dots, Y_{i,m})'$, $i = 1, \dots, n$, with finite dimensional unknown parameter $(\Lambda_1, \dots, \Lambda_m, \delta, r, \theta)'$, where $\Lambda_i = \Lambda(t_i)$. Maximum likelihood estimation of t distribution was discussed by many authors in the literature, see e.g. [17–21]. When the observation times become random and different for each subject, this problem is changed to an infinite dimensional statistical problem because the parameter space is now infinite dimension and we need estimate a function $\Lambda(t)$. However, due to the correlation between the successive degradation measurements and monotonicity of $\Lambda(t)$, we need to solve a difficult constrained optimization problem. Both algorithms and asymptotics pose many challenges. In the following, we first derive the large sample properties of the MLE, then use EM algorithm to obtain the MLE.

3. Maximum likelihood estimator

Suppose we observe the degradation process $Y(t)$ at a random number K of random times $0 = T_{K,0} < T_{K,1} < \cdots < T_{K,K}$. Represent $\underline{T}_K = (T_{K,1}, \dots, T_{K,K})$ and $\underline{Y}_K = (Y_{K,1}, \dots, Y_{K,K})$, where $Y_{K,j} = Y(T_{K,j})$. Assume $(K, \underline{T}_K)$ is independent of $Y(t)$. Suppose we observe n i.i.d. copies of $X = (\underline{Y}_K, \underline{T}_K, K), X_1, \dots, X_n$, where $X_i = (\underline{Y}_{K_i}^{(i)}, \underline{T}_{K_i}^{(i)}, K_i)$ for $i = 1, \dots, n$. Our goal is to estimate the unknown parameters of the process, $(\Lambda(t), r, \delta, \theta)$.

The log likelihood function is given by

$$\begin{aligned} l_n(\Lambda, r, \delta, \theta) = \sum_{i=1}^{n} \Bigg\{ & \log\Gamma\left(\delta + \frac{K_i}{2}\right) - \log\Gamma(\delta) - \frac{K_i}{2}\log r - \frac{1}{2}\log|A_i| \\ & - \left(\delta + \frac{K_i}{2}\right)\log\left[1 + \frac{1}{2r}(y^{(i)} - \lambda^{(i)})'A_i^{-1}(y^{(i)} - \lambda^{(i)})\right]\Bigg\}, \end{aligned} \tag{9}$$

where $y^{(i)} = (\Delta Y_{K_i,1}^{(i)}, \dots, \Delta Y_{K_i,K_i}^{(i)})'$, $\lambda^{(i)} = (\Delta\Lambda_{K_i,1}, \dots, \Delta\Lambda_{K_i,K_i})'$, $\Delta Y_{K_i,j}^{(i)} = Y_{K_i,j}^{(i)} - Y_{K_i,j-1}^{(i)}$, $\Delta\Lambda_{K_i,j} = \Lambda(T_{K_i,j}^{(i)}) - \Lambda(T_{K_i,j-1}^{(i)})$, and the elements of A_i are given by $[A_i]_{kk} = (\Delta\Lambda_{K_i,k})^2\theta + \Delta\Lambda_{K_i,k}$ and $[A_i]_{kj} = \Delta\Lambda_{K_i,k}\Delta\Lambda_{K_i,j}\theta$ for $k \neq j$. Let $(\hat{\Lambda}_n, \hat{r}_n, \hat{\delta}_n, \hat{\theta}_n)$ be the MLE and then

$$(\hat{\Lambda}_n, \hat{r}_n, \hat{\delta}_n, \hat{\theta}_n) = \arg\max_{\mathcal{F}\times\mathcal{R}_+^3} l_n(\Lambda, r, \delta, \theta)$$

where $\mathcal{R}^+ \subset (0, \infty)$ is a compact interval and

$$\mathcal{F} = \{\Lambda : [0, \infty) \to [0, \infty) | \Lambda \text{ is a nondecreasing function with } \Lambda(0) = 0\}.$$

Let $C_t = \{T_{K_i,j}^{(i)}, j = 1, \dots, K_i, i = 1, \dots, n\} = \{T_1, \dots, T_N\}$ be the superset of all the inspection times for all subjects. The MLE $\hat{\Lambda}_n(t)$ can only be identified at $T_1, \dots, T_N$ and $\hat{\Lambda}_n$ can be defined as a nondecreasing piecewise linear function with possible knots at $T_1, \ldots, T_N$. The choice of making $\hat{\Lambda}_n(t)$ a piecewise linear function is arbitrary and other conventions are possible (e.g., making $\hat{\Lambda}_n(t)$ as a step function with possible jumps at $T_1, \ldots, T_N$).

Characterizing $\hat{\alpha}_n = (\hat{\Lambda}_n, \hat{r}_n, \hat{\delta}_n, \hat{\theta}_n)$ can be easily formulated by the following optimality condition (e.g. [22]). Write $\mathcal{F}$ as $\{\underline{\Lambda} \in R^N : 0 \le \Lambda_1 \le \Lambda_2 \le \cdots \le \Lambda_N\}$ and denote

$$\phi_i(\alpha) = \frac{\partial l_n(\alpha)}{\partial \Lambda_i}, \quad i = 1, \ldots, N,$$

and

$$\phi_r(\alpha) = \frac{\partial l_n(\alpha)}{\partial r}, \qquad \phi_\delta(\alpha) = \frac{\partial l_n(\alpha)}{\partial \delta}, \qquad \phi_\theta(\alpha) = \frac{\partial l_n(\alpha)}{\partial \theta},$$

which can be calculated explicitly from (9). If $\hat{\alpha}_n = \arg\max_{\mathcal{F}\times\mathcal{R}_+^3} l_n(\alpha)$, then $\nabla l_n(\hat{\alpha}_n)'(\alpha - \hat{\alpha}_n) \le 0$ for all $\alpha \in \mathcal{F} \times \mathcal{R}_+^3$, i.e.,

$$\sum_{i=1}^{N} \phi_i(\hat{\alpha}_n)(\Lambda_i - \hat{\Lambda}_{n,i}) + \phi_r(\hat{\alpha}_n)(r - \hat{r}_n) + \phi_\delta(\hat{\alpha}_n)(\delta - \hat{\delta}_n) + \phi_\theta(\hat{\alpha}_n)(\theta - \hat{\theta}_n) \le 0, \tag{10}$$

for all $(\Lambda_1, \ldots, \Lambda_N, r, \delta, \theta) \in \mathcal{F} \times \mathcal{R}_+^3$.

In the following, we study the asymptotic properties of the MLE. Let $\mathcal{B}$ denote the collection of Borel sets in $\mathbb{R}$. Let $\mathcal{B}_{[0,T]} = \{B \cap [0, T] : B \in \mathcal{B}\}$ for some fixed constant T. On $([0, T], \mathcal{B}_{[0,T]})$ we define measures μ, γ as follows: for $B_1, B_2 \in \mathcal{B}_{[0,T]}$, define

$$\mu(B_1 \times B_2) = \sum_{k=1}^{\infty} P(K = k) \sum_{j=1}^{k} P(T_{K,j-1} \in B_1, T_{K,j} \in B_2 | K = k),$$

$$\gamma(B) = \sum_{k=1}^{\infty} P(K = k) P(T_{k,k} \in B | K = k).$$

Based on measure μ, for any $\alpha_1 = (\Lambda_1, r_1, \delta_1, \theta_1)$ and $\alpha_2 = (\Lambda_2, r_2, \delta_2, \theta_2)$, define the L_2 metric $d(\alpha_1, \alpha_2)$ as

$$\begin{aligned} d^2(\alpha_1, \alpha_2) &= (r_1 - r_2)^2 + (\delta_1 - \delta_2)^2 + (\theta_1 - \theta_2)^2 + \|\Delta\Lambda_1 - \Delta\Lambda_2\|_\mu^2 \\ &= (r_1 - r_2)^2 + (\delta_1 - \delta_2)^2 + (\theta_1 - \theta_2)^2 + \int [(\Lambda_1(v) - \Lambda_1(u)) - (\Lambda_2(v) - \Lambda_2(u))]^2 \mathrm{d}\mu(v, u). \end{aligned}$$

Wellner and Zhang [23] showed that this metric d is equivalent to another common L_2 metric d_0: If $P(K \le k_0) = 1$ for some $k_0 < \infty$, $\frac{1}{2} d(\alpha_1, \alpha_2) \le d_0(\alpha_1, \alpha_2) \le k_0 d(\theta_1, \theta_2)$, where

$$\begin{aligned} d_0^2(\theta_1, \theta_2) &= (r_1 - r_2)^2 + (\delta_1 - \delta_2)^2 + (\theta_1 - \theta_2)^2 + \|\Lambda_1 - \Lambda_2\|_{\mu_0}^2 \\ &= (r_1 - r_2)^2 + (\delta_1 - \delta_2)^2 + (\theta_1 - \theta_2)^2 + \int (\Lambda_1(t) - \Lambda_2(t))^2 \mathrm{d}\mu_0(t), \end{aligned}$$

and $\mu_0(B) = \sum_{k=1}^{\infty} P(K = k) \sum_{j=1}^{k} P(T_{K,j} \in B | K = k)$.

Let $(\Lambda_0, r_0, \delta_0, \theta_0)$ be the true parameters. To establish consistency of the MLE, assume the following regularity conditions:

Condition A1. The true parameter $(r_0, \delta_0, \theta_0)$ is in the interior of $\mathcal{R}_+^3$.

Condition A2. The observation times $T_{K,j}, j = 1, \ldots, K$ are random and take values in $[0, T]$ with $T < \infty$.

Condition A3. $E(K) < \infty$ and $\Lambda_0(T) < \infty$.

The above conditions are generally mild conditions in the context of applications. Conditions A1–A3 usually hold in practice.

Theorem 3.1 (*Consistency*). *Suppose that the conditions* A1–A3 *hold. Then, for every $t < T$ and $\gamma([t, T]) > 0$,*

$$d((\hat{\Lambda}_n 1_{[0,t]}, \hat{r}_n, \hat{\delta}_n, \hat{\theta}_n), (\Lambda_0 1_{[0,t]}, r_0, \delta_0, \theta_0)) \to 0, \tag{11}$$

almost surely as $n \to \infty$. If $\gamma(\{T\}) > 0$, we also have

$$d((\hat{\Lambda}_n, \hat{r}_n, \hat{\delta}_n, \hat{\theta}_n), (\Lambda_0, r_0, \delta_0, \theta_0)) \to 0.$$

To obtain the convergence rate of the MLE, we also assume that:

Condition B1. For some interval $[T_l, T_u]$ with $T_l > 0$ and $\Lambda_0(T_l) > 0$, $P(\cap_{j=1}^{K} T_{K,j} \in [T_l, T_u]) = 1$.

Condition B2. $P(K < k_0) = 1$ for some $k_0 < \infty$.

Condition B3. There exists a constant $c > 0$ such that $P(T_{K,j} - T_{K,j-1} \ge c$ for all $j = 1, \ldots, K) = 1$.

Condition B4. Function Λ_0 is differentiable and there exist constants b_l and b_u such that $0 < b_l < \Lambda_0'(t) < b_u < \infty$.

Condition B1 assumes $T_{K,j}$ is bounded away from zero and Condition B2 assumes K is finite almost surely. Condition B3 assures that observation time points are separated and Condition B4 assumes that there is no flat part in $\Lambda_0(t)$ and its derivative is bounded away from zero and infinity.

Theorem 3.2 (*Convergency Rate*). *Suppose the conditions* A1–A4 *and* B1–B4 *hold. Then,*

$$n^{1/3}d((\hat{\Lambda}_n, \hat{r}_n, \hat{\delta}_n, \hat{\theta}_n), (\Lambda_0, r_0, \delta_0, \theta_0)) = O_p(1). \tag{12}$$

The overall convergence rate for the MLE is $n^{1/3}$. Simulation study in Section 5 suggests that the convergence rate of $(\hat{r}_n, \hat{\delta}_n, \hat{\theta}_n)$ of the finite dimensional parameters is still $n^{1/2}$. The proofs of Theorems 3.1 and 3.2 are given in Appendix which are established similarly as in [23,24]).

In the following, bootstrap method (e.g. [25]) is used to assess the variability of the MLE: the degradation paths are resampled independently, i.e., we resample triples $X_i^* = (\underline{Y}_{K_i}^{(i)*}, \underline{T}_{K_i}^{(i)*}, K_i^*)$, $i = 1, \ldots, n$, i.i.d. from the empirical distribution putting mass n^{-1} to $X_i = (\underline{Y}_{K_i}^{(i)}, \underline{T}_{K_i}^{(i)}, K_i)$, $i = 1, \ldots, n$, and define

$$(\hat{\Lambda}_n^*, \hat{r}_n^*, \hat{\delta}_n^*, \hat{\theta}_n^*) = \arg\max_{\mathcal{F}\times\mathcal{R}_+^3} l_n(\Lambda, r, \delta, \theta; X_1^*, \ldots, X_n^*).$$

Theorem 3.3. (i). *Under the same conditions of Theorem 3.1, for every $t < T$ and $\gamma([t, T]) > 0$, $d((\hat{\Lambda}_n^* 1_{[0,t]}, \hat{r}_n^*, \hat{\delta}_n^*, \hat{\theta}_n^*), (\hat{\Lambda}_n 1_{[0,t]}, \hat{r}_n, \hat{\delta}_n, \hat{\theta}_n)) \to 0$, almost surely as $n \to \infty$. If $\gamma(\{T\}) > 0$, we also have $d((\hat{\Lambda}_n^*, \hat{r}_n^*, \hat{\delta}_n^*, \hat{\theta}_n^*), (\hat{\Lambda}_n 1_{[0,t]}, \hat{r}_n, \hat{\delta}_n, \hat{\theta}_n)) \to 0$.*

(ii). *Under the same conditions of Theorem 3.2, $n^{1/3}d((\hat{\Lambda}_n^*, \hat{r}_n^*, \hat{\delta}_n^*, \hat{\theta}_n^*), (\hat{\Lambda}_n, \hat{r}_n, \hat{\delta}_n, \hat{\theta}_n)) = O_p(1)$.*

Theorem 3.3 shows that the bootstrap estimator is consistent with $n^{1/3}$ convergence rate. To construct confidence sets for unknown parameters, we use both bootstrap-percentile method [26] and bootstrap-t method [27]. Let $\hat{\omega}_n$ be an estimator of one of the unknown parameters, r, δ, θ, or $\Lambda(t)$ for fixed t and $\hat{\omega}_n^*$ be its bootstrap analog based on $X_1^*, \ldots, X_n^*$. Define $H_{\text{boot}}(x) = P(\hat{\omega}_n^* \le x | X_1^*, \ldots, X_n^*)$. The bootstrap $(1 - 2\alpha)$ percentile confidence set for ω is $[H_{\text{boot}}^{-1}(\alpha), H_{\text{boot}}^{-1}(1 - \alpha)]$. The bootstrap-$t$ method is based on a given studentized "pivot" $\mathcal{R}_n = (\hat{\omega}_n - \omega_0)/\hat{\sigma}_{\omega,n}$, where $\hat{\sigma}_{\omega,n}$ is the variance estimator of $\hat{\omega}_n$. The distribution G_n of $\mathcal{R}_n$ is usually unknown and it is estimated by the bootstrap estimator G_{boot} defined by $G_{\text{boot}}(x) = P(\mathcal{R}_n^* \le x | X_1, \ldots, X_n)$, where $\mathcal{R}_n^* = (\hat{\omega}_n^* - \hat{\omega}_n)/\hat{\sigma}_{\omega,n}^*$, and $\hat{\omega}_n^*$ and $\hat{\sigma}_{\omega,n}^*$ are bootstrap analogs of $\hat{\omega}_n$ and $\hat{\sigma}_{\omega,n}$, respectively. The resulting $(1 - 2\alpha)$-confidence set for ω is $[\hat{\omega}_n - \hat{\sigma}_{\omega,n}G_{\text{boot}}^{-1}(\alpha), \hat{\omega}_n - \hat{\sigma}_{\omega,n}G_{\text{boot}}^{-1}(1 - \alpha)]$. Here, we use the bootstrap variance estimator $v_{\text{boot},\omega}$ for $\hat{\sigma}_{\omega,n}$. Since both G_{boot} and $v_{\text{boot},\omega}$ are approximated by Monte Carlo, we need to do nested bootstrapping. We first generate B_1 bootstrap data sets to approximate G_{boot} and for each given bootstrap data set, we generate B_2 bootstrap date sets to approximate $v_{\text{boot},\omega}$. Simulation studies in Section 5 will display the performance of these bootstrap estimators based on coverage probabilities and interval lengths.

4. Computation of the MLE

The difficulty to obtain the MLE is that we have to maximize the log likelihood function (9) under the order restriction that function $\Lambda(t)$ is monotone. Direct constrained optimization of the likelihood function is computationally difficult, especially when the inspection times for different units are different. In the following, we use EM algorithm to find the MLE iteratively.

Let $C_t = \{T_1, \ldots, T_N\}$ be the superset of all the inspection times for all subjects. For each subject i, we have degradation observations at only a subset of these time points. The steps to get the MLE are that we first impute the sufficient statistics such as the unobserved degradation measurements at the unobserved time points and the random effects given the observed data at other points, current values of parameters. Once we do this, we are in a special case where we can get the MLE explicitly.

When both $\{\underline{Y}_N^{(i)}, i = 1 \ldots, n\}$ and $\{\nu_i, \omega_i, i = 1, \ldots, n\}$ are considered observed, $\{\underline{Y}_N^{(1)}, \ldots, \underline{Y}_N^{(n)}, \nu_1, \ldots, \nu_n, \omega_1, \ldots, \omega_n\}$ comprise the complete data. Because of the conditional structure of the complete data model given by (2) and (3), the complete data likelihood function can be factored into the product of two distinct functions. Hence, given the complete data, the log likelihood function of the parameters $(\Lambda, r, \delta, \theta)$ is given by

$$l(\Lambda, r, \delta, \theta) = l_N(\Lambda) + l_G(r, \delta, \theta), \tag{13}$$

where

$$l_N(\Lambda) = \sum_{i=1}^{n}\left[\frac{N}{2}\log\omega_i - \frac{1}{2}\sum_{j=1}^{N}\log\lambda_j - \frac{\omega_i}{2}\sum_{j=1}^{N}\frac{(y_{ij} - \nu_i\lambda_j)^2}{\lambda_j}\right] \tag{14}$$

and

$$l_G(r, \delta, \theta) = \sum_{i=1}^{n}\left[\left(\delta - \frac{1}{2}\right)\log\omega_i - \frac{1}{2}\log\theta - \frac{\omega_i(\nu_i - 1)^2}{2\theta} + \delta\log r - r\omega_i - \log\Gamma(\delta)\right], \tag{15}$$

where $y_{ij} = Y^{(i)}_{N,j} - Y^{(i)}_{N,j-1}$ and $\lambda_j = \Lambda(T_j) - \Lambda(T_{j-1})$, $i = 1, \dots, n$, $j = 1, \dots, N$. The complete data sufficient statistics for $(\Lambda, r, \delta, \theta)$ are

$$\sum_{i=1}^{n} \omega_i \nu_i^2, \qquad \sum_{i=1}^{n} \omega_i \nu_i, \qquad \sum_{i=1}^{n} \omega_i, \qquad \sum_{i=1}^{n} \log \omega_i, \qquad \sum_{i=1}^{n} \omega_i y_{ij}^2, \quad j = 1, \dots, N.$$

Given the sufficient statistics, the MLE of $\Lambda(t)$ and the MLE of (r, δ, θ) can be obtained from $l_N(\Lambda)$ and $l_G(r, \delta, \theta)$, respectively. From (14) and (15), we have

$$\hat{\lambda}_j = \frac{-n + \sqrt{n^2 + 4\sum\limits_{i=1}^{n} w_i \nu_i^2 \sum\limits_{i=1}^{n} w_i y_{ij}^2}}{2\sum\limits_{i=1}^{n} w_i \nu_i^2}, \quad j = 1, \dots, N, \tag{16}$$

which are positive. So, $\hat{\Lambda}(T_j) = \sum_{k=1}^{j} \hat{\lambda}_k$ gives the maximum likelihood estimate of $\Lambda(t)$. The maximum likelihood estimates of (r, δ, θ) from $l_G(\Lambda, r, \delta, \theta)$ are

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} \omega_i (\nu_i - 1)^2, \qquad \hat{r} = \frac{\hat{\delta}}{\sum\limits_{i=1}^{n} \omega_i / n} \tag{17}$$

and $\hat{\delta}$ is the solution of

$$\psi(\delta) - \log \delta = \frac{1}{n}\sum_{i=1}^{n} \log \omega_i - \log\left[\frac{1}{n}\sum_{i=1}^{n} \omega_i\right], \tag{18}$$

where ψ is the digamma function and $\psi(x) - \log(x)$ is an increasing function with limits $-\infty$ and zero as x goes to zero and infinity respectively [20]. Thus, Eq. (18) has a unique zero $\hat{\delta}$.

Let $Y_{i,\text{obs}}$ denote the observed components for unit i and $Y_{i,\text{miss}}$ denote the missing components of unit i. Also, let $\Omega_t = \{\Lambda^t, r^t, \delta^t, \theta^t\}$ be the current values of the parameters. Since $(Y_{i,\text{obs}}, Y_{i,\text{miss}})$ follows multivariate t distribution with scatter matrix, say Σ, it is well known that the conditional distribution of $Y_{i,\text{miss}}$ given $Y_{i,\text{obs}}$ again is multivariate t distributed and

$$E(Y_{i,\text{miss}} | Y_{i,\text{obs}}, \Omega_t) = \eta_1 - I_{11}^{-1} I_{12} (Y_{i,\text{obs}} - \eta_2), \tag{19}$$

where $\Sigma^{-1} = [I_{11}, I_{12}; I_{21}, I_{22}]$ and η_1 and η_2 are the mean vectors of $Y_{i,\text{miss}}$ and $Y_{i,\text{obs}}$ respectively. This can be used to get imputed values for the unobserved degradation data at the unobserved inspection times. For subject i, the conditional distribution of ν_i given Ω_t is a noncentral t distribution, the conditional distribution of ν_i given Ω_t and ω_i is a normal distribution, and the conditional distribution of w_i given Ω_t follows a Gamma distribution. Thus for the expectations of the sufficient statistics at the $(t+1)$th E-step, we have

$$E[w_i | \Omega_t] = \left(\delta + \frac{K_i}{2}\right)\left[-\frac{(\theta^{-1} + \sum\limits_{j=1}^{K_i} y_{ij})^2}{2(\theta^{-1} + \sum\limits_{j=1}^{K_i} \lambda_{ij})} + \sum_{j=1}^{K_i} \frac{y_{ij}^2}{2\lambda_{ij}} + \frac{1}{2\theta} + r\right]^{-1}, \tag{20}$$

$$E[\log w_i | \Omega_t] = -\log\left[-\frac{(\theta^{-1} + \sum\limits_{j=1}^{K_i} y_{ij})^2}{2(\theta^{-1} + \sum\limits_{j=1}^{K_i} \lambda_{ij})} + \sum_{j=1}^{K_i} \frac{y_{ij}^2}{2\lambda_{ij}} + \frac{1}{2\theta} + r\right] + \psi\left(\delta + \frac{K_i}{2}\right), \tag{21}$$

$$E[\omega_i \nu_i | \Omega_t] = E[\omega_i E(\nu_i | \omega_i, \Omega_t) | \Omega_t] = E[w_i | \Omega_t] \frac{\theta^{-1} + \sum\limits_{j=1}^{K_i} y_{ij}}{\theta^{-1} + \sum\limits_{j=1}^{K_i} \lambda_{ij}}, \tag{22}$$

$$E[\omega_i \nu_i^2 | \Omega_t] = E[\omega_i E(\nu_i^2 | \omega_i, \Omega_t) | \Omega_t] = E[w_i | \Omega_t] \left[\frac{\theta^{-1} + \sum\limits_{j=1}^{K_i} y_{ij}}{\theta^{-1} + \sum\limits_{j=1}^{K_i} \lambda_{ij}}\right]^2 + \left(\sum_{j=1}^{K_i} \lambda_{ij} + \theta^{-1}\right)^{-1}. \tag{23}$$

Table 1
Results of Monte Carlo study for (r, δ, θ) estimates based on 1000 repeated samples for data generated from conditional Wiener process.

		$n = 50$			$n = 100$		
		$\hat{r}_n$	$\hat{\delta}_n$	$\hat{\theta}_n$	$\hat{r}_n$	$\hat{\delta}_n$	$\hat{\theta}_n$
$\Lambda(t) = t$	BIAS	0.2766	0.0236	−0.0161	0.1939	0.0146	−0.0094
	SD	1.3547	0.1836	0.2391	0.9321	0.1302	0.1700
	BOOT-P-CP	0.932	0.905	0.924	0.936	9.957	0.940
	BOOT-T-CP	0.959	0.935	0.943	0.958	9.942	0.948
$\Lambda(t) = t^2/10$	BIAS	0.3553	0.0346	0.0236	0.3074	0.0122	0.0179
	SD	1.4320	0.1889	0.2546	0.9903	0.1333	0.1795
	BOOT-P-CP	0.933	0.921	0.924	0.935	9.939	0.937
	BOOT-T-CP	0.941	0.953	0.942	0.941	9.948	0.942
$\Lambda(t) = \sqrt{10t}$	BIAS	0.3764	0.0409	−0.0008	0.2845	0.0361	−0.0005
	SD	1.2668	0.1801	0.2530	0.9003	0.1265	0.1801
	BOOT-P-CP	0.918	0.903	0.923	0.925	9.916	0.944
	BOOT-T-CP	0.942	0.958	0.928	0.949	9.948	0.946

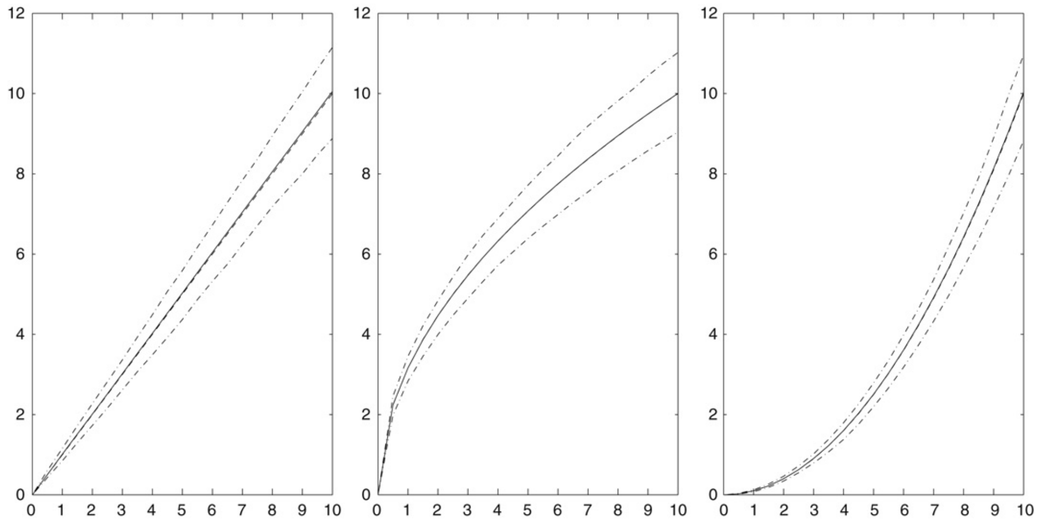

Fig. 1. The 95% pointwise confidence interval for the $\Lambda(t)$ when $n = 50$. The dashed lines are true functions, the solid line is the mean of the MLEs from 1000 simulations, and the dash-dotted is the 95% pointwise confidence interval for the $\Lambda(t)$.

The complete algorithm consists of the following steps.
Step 1. Choose initials $\Lambda^0(t), \theta^0, \delta^0$ and r^0.
Step 2. For given $(\Lambda^p(t), \theta^p, \delta^p, r^p)$ $(p = 0, 1, 2, \ldots)$, impute the sufficient statistics by (19)–(23).
Step 3. Update Λ^{p+1} by (16) and update $(\theta^{p+1}, \delta^{p+1}, r^{p+1})$ by (17) and (18).
Step 4. Repeat Steps 2–3 until the change of the likelihood function within pre-specified threshold.

5. Simulation

Let $\{\underline{Y}_{K_i}, \underline{T}_{K_i}, K_i\}$, $i = 1, \ldots, n$ be a random sample. We choose $K_i \in \{18, 19, 20, 21, 22\}$ and $P(K_i = k) = 1/5$ for $k = 18, \ldots, 22$. Then $\underline{T}_{K_i}$ are made from the order statistics of K_i random observations from uniform(0,10). The time points are rounded to the first decimal point to make the observation times possibly tied. The degradation measurements $\underline{Y}_{K_i}$ are generated from conditional Wiener process, that is,

$$\omega_i \sim \text{Gamma}(\delta, r^{-1}), \qquad \nu_i|\omega_i \sim N(1, \theta/\omega_i),$$
$$Y_{K_i,j} - Y_{K_i,j-1}|(\nu_i, \omega_i) \sim N\left[\nu_i(\Lambda(T_{K_i,j}) - \Lambda(T_{K_i,j-1})), \omega_i^{-1}(\Lambda(T_{K_i,j}) - \Lambda(T_{K_i,j-1}))\right],$$

where $r = .5$, $\delta = 4$, $\theta = 1$ and $\Lambda(t)$ is chosen as one of the three functions with different shapes, $t, t^2/10, \sqrt{10t}$. We choose the number of subjects $n = 50$ and 100, respectively. We carry out a Monte Carlo study by repeating the simulation 1000 times. For each simulated degradation dataset, the bootstrap Monte Carlo size is 400. When we need the bootstrap variance estimator to construct confidence set, for each bootstrap data, we generate 200 bootstrap datasets to approximate variance estimator.

The bias and standard error for the MLE of (r, δ, θ) are given in Table 1. The table shows that the sample bias for $(\hat{r}_n, \hat{\delta}_n, \hat{\theta}_n)$ is small. The ratio of the standard error for the two sample sizes are close to $\sqrt{2}$, which indicates the convergence rate

Table 2
The coverage probabilities (CP) and the expected length ratios (LR) of the 95% pointwise confidence intervals for $\Lambda(t)$ when $\Lambda(t) = t$.

t	$n = 50$				$n = 100$			
	BOOT-P		BOOT-T		BOOT-P		BOOT-T	
	CP	LR	CP	LR	CP	LR	CP	LR
1	0.956	1.1644	0.952	1.0923	0.944	1.0917	0.950	1.1841
2	0.967	1.1561	0.948	1.1352	0.936	0.1492	0.946	1.1232
4	0.976	1.2338	0.955	1.1225	0.955	1.0896	0.952	1.0329
6	0.971	1.2655	0.963	1.1776	0.963	0.9474	0.958	1.0310
8	0.983	1.3169	0.969	1.1907	0.969	0.9758	0.961	1.0161
9	0.962	1.3038	0.960	1.2038	0.960	1.0018	0.959	0.9985

of $(\hat{r}_n, \hat{\delta}_n, \hat{\theta}_n)$ is actually $\sqrt{n}$. The 95% bootstrap-percentile confidence interval coverage probability (BOOT-P-CP) and 95% bootstrap-t confidence interval coverage probability (BOOT-T-CP) are also showed in Table 1. Both coverage probabilities are close to the nominal value but the bootstrap-t confidence interval has more accurate CP.

Fig. 1 displays the 95% pointwise confidence intervals for the $\Lambda(t)$ based on these 1000 simulations, along with three different true functions when $n = 50$. The dashed lines are true functions, the solid line is the mean of the MLEs from 1000 simulations, and the dash-dotted is the 95% pointwise confidence interval. Apparently, the MLE is close and converges to the true function. To assess the performance of the bootstrap method, Table 2 gives the coverage probabilities of the bootstrap-percentile and bootstrap-t confidence intervals for $\Lambda(t)$ when $t = 1, 2, 4, 6, 8, 9$ when $\Lambda(t) = t$. Table 2 also shows the confidence interval length ratios for these two types of bootstrap confidence intervals, where the ratio is defined as expected bootstrap confidence interval length over the confidence interval length obtained from 1000 simulations directly showed in Fig. 1. Both types of the confidence interval perform reasonably good jobs, where the coverage probability is close to 0.95 and the length ratio is close to 1.

6. Application: bridge beam data

Elsayed and Liao [4] presented the data about the degradation of the bridge beams due to chloride ion ingression. A sample of 20 bridge beams is observed. The field data considered are bivariate data (t_j, y_{ij}), in which y_{ij} is the measurement of the loss of strength for bridge beam i after j years, $i = 1, \ldots, 20, j = 10, \ldots, 40$. The left panel of Fig. 2 shows the bridge beams' strength losses from 10 to 40 years. Elsayed and Liao [4] used the data to illustrate statistical methodology by fitting integrated geometric Brownian motion model. This model provides a good fit, however, it does not have tractable forms for both the joint density of degradation measurements and the first passage time distribution.

We reconsider the same data to illustrate the application of nonhomogeneous Wiener process model with random effects. For the discussion here we assume failure occurs when the loss of strength exceeds 400 pst. On fitting the model (2) and (3), maximum likelihood estimates were obtained as follows:

$$\hat{\delta} = 2.48([0.11, 5.96]), \qquad \hat{r} = 0.0013([0.0000, 0.0034]), \qquad \hat{\theta} = 200.20([125.48, 345.47]).$$

In parentheses give the 95% confidence intervals for the unknown parameters by bootstrap-t method. Small value of $\hat{r}$ and large value of $\hat{\theta}$ indicate strong evidence of existing of the random effects. One reason to explain this is that the original data have smoother sample paths than seen with Wiener processes. The right panel of Fig. 2 (solid line) is the nonparametric estimate of $\Lambda(t)$. The dashed lines are 95% pointwise confidence band for $\Lambda(t)$. The variabilities of the MLEs are obtained by bootstrap-t method.

In Fig. 3, for an illustrative set of units ($i = 3, 7, 10, 14, 20$), the observed degradation paths $Y_i(t)$ together with the fitted values which were computed from $\hat{\Lambda}_i(t) = \hat{\nu}_i \hat{\Lambda}(t)$, where $\hat{\nu}_i$ is the estimated random effect for unit i computed from (20). As we can see, the Wiener process model with random effects fits the data quite well. The TTF distribution function $F(t)$ is easily estimated by inserting parameter estimates in (5). The MLE of $F(t)$ is given by the left panel of Fig. 4, along with the 95% pointwise confidence band computed by Bootstrap method. The right panel of Fig. 4 plots the Kaplan–Meier estimated, computed from follow-up of each unit until it crosses the failure threshold. The estimated curve, based on nonhomogeneous Wiener process with random effects appears to follow the data very well.

7. Conclusion

We develop a nonhomogeneous Wiener process model with random effects for degradation analysis. The probability distribution of sample path measurements at discrete time points has a simple closed form, as does the time to failure distribution. This model defines a wide class of time to failure distributions with different choices of the degradation function. Consistency and convergence rate are established for the MLE. Current research is investigating the asymptotic distribution of the MLE.

The imperfect procedures and equipment can produce measurement errors in the degradation experiment. An extension of our model, similar to that described in [10], can take measurement errors into account. We may add independent errors

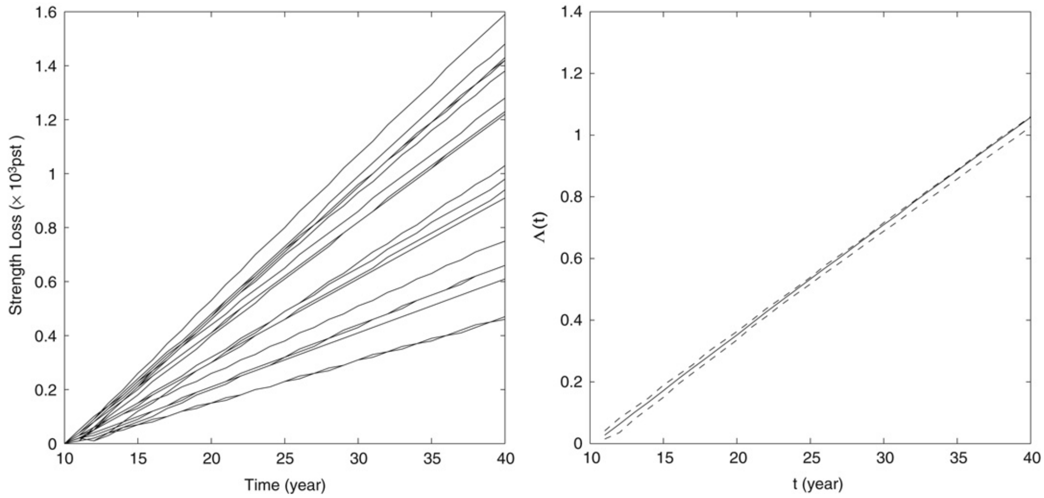

Fig. 2. Bridge beams' strength loss over time.

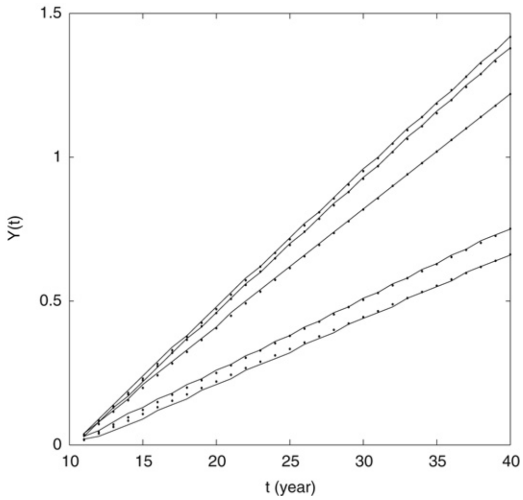

Fig. 3. The fitted curves for units 3, 7, 10, 14, 20.

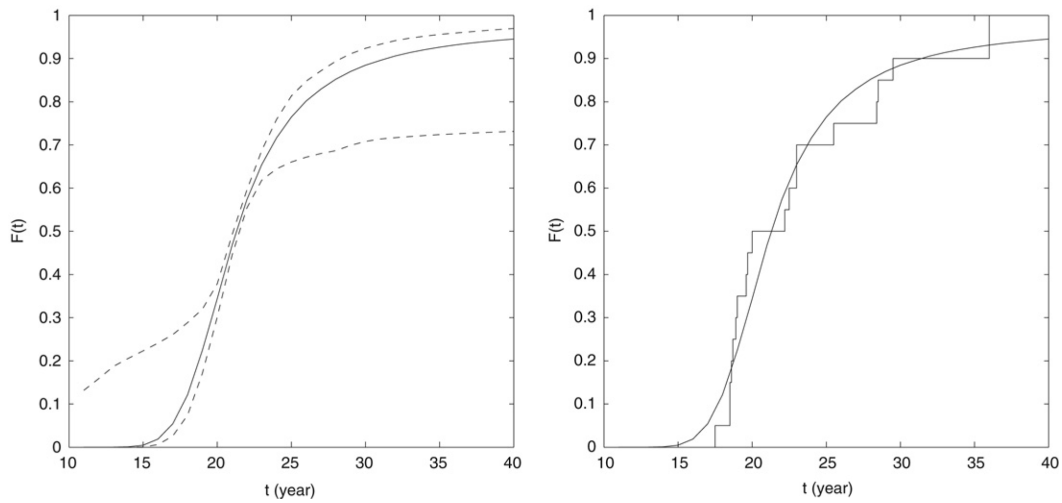

Fig. 4. Estimates of the failure time distribution function $F(t)$.

or intercorrelated measurement errors since readings on test items made at the same time under the same test conditions can be highly correlated. For example, assume the degradation measurements at $t_1 < t_2 < \cdots < t_n$ are given by

$$Y(t_i) = \Lambda(t_i) + \sigma_1 W(\Lambda(t_i)) + \sigma_2\epsilon_i, \tag{24}$$

where ϵ_i denotes the measurement error and is assumed to be distributed as N(0, 1). Parameter inference of this model is straightforward, but the nonparametric estimation need additional research. Note that the increments of (24) are not independent anymore. Even when the measurement times are the same for all subjects, there is no closed form for the MLE. After incorporating the random effects into the model, the likelihood function gets much more complicated. There are many challenges for both finding efficient algorithms and deriving large sample properties for the estimators.

Another interesting extension for current model is when we have covariate information. For example, in accelerated degradation experiment, the stress level such as temperature is the covariate. Bagdonavicius and Nikulin [28] incorporated the covariates by replacing $\Lambda(t)$ by $\Lambda(te^{x^{\mathrm{T}}\beta})$, where x is the covariate vector. Lawless and Crowder [16] treated scale parameter of the Gamma process as a function of x to accommodate the covariate. Similar to Cox model, we may study the proportional mean model replacing $\Lambda(t)$ by $\Lambda(t)e^{x^{\mathrm{T}}\beta}$ to incorporate the covariate information. For all these models, we can study the likelihood inference in a similar way.

Acknowledgments

This research is supported by the US National Science Foundation. The author would like to thank the guest editor Professor Frederic Ferraty and two referees for providing valuable comments to improve the original manuscript. The author would like to thank Professor Vijay Nair for many helpful discussions and suggestions.

Appendix

Let $\mathbb{P}_n$ denote the empirical measure and $\mathbb{G}_n$ denote the empirical process. Let C or C_i, $i = 1, 2, \ldots,$ stand for generic constants which may change from line to line in the proof. Denote $\alpha = (\Lambda, r, \delta, \theta)$.

Proof of Theorem 3.1. Let $M_n(\alpha) = n^{-1}l_n(\Lambda, r, \delta, \theta) = \mathbb{P}_n m_\alpha(X)$ and $M(\alpha) = Pm_\alpha(X)$, where

$$m_\alpha(X) = \log\left[\frac{\Gamma\left(\delta + \frac{K}{2}\right)}{\Gamma(\delta)}\right] - \frac{K}{2}\log r - \frac{1}{2}\log|A| - \left(\delta + \frac{K}{2}\right)\log\left[1 + \frac{1}{2r}(y-\lambda)'A^{-1}(y-\lambda)\right].$$

We first show that $\hat\alpha_n = (\hat\Lambda_n, \hat r_n, \hat\delta_n, \hat\theta_n)$ is uniformly bounded. Note that $(\hat r_n, \hat\delta_n, \hat\theta_n)$ is bounded since they are in a bounded compact set $\mathcal{R}_+^3$. Let $\tilde\alpha_n = (\hat\Lambda_n + \epsilon\Lambda, \hat r_n, \hat\delta_n, \hat\theta_n)$ for any given ϵ and Λ. Since $M_n(\tilde\alpha) \le M_n(\hat\alpha)$, it follows that

$$\begin{aligned}
0 &\ge \lim_{\epsilon\downarrow 0}\frac{1}{\epsilon}[M_n(\tilde\alpha) - M_n(\hat\alpha)] \\
&= \mathbb{P}_n\left\{-\frac{1}{2}\left[\sum_{j=1}^K \frac{\lambda_{Kj}}{\hat\lambda_{nKj}} + \frac{\sum_{j=1}^K \lambda_{Kj}}{\hat\theta_n^{-1} + \sum_{j=1}^K \hat\lambda_{nKj}}\right] - \frac{\hat\delta_n + \frac{K}{2}}{2\hat r_n + (y-\hat\lambda)'\hat A^{-1}(y-\hat\lambda)}\right. \\
&\quad \left.\times\left[-\frac{\sum_{j=1}^K \lambda_{Kj}\left(\sum_{j=1}^K y_{Kj} + \hat\theta_n^{-1}\right)^2}{\left(\hat\theta_n^{-1} + \sum_{j=1}^K \hat\lambda_{nKj}\right)^2} + \sum_{j=1}^K \frac{y_{Kj}^2\lambda_{Kj}}{\hat\lambda_{nKj}^2}\right]\right\}.
\end{aligned}$$

Letting $\Lambda = -\hat\Lambda_n$ and after a straightforward algebra, this yields $\mathbb{P}_n\sum_{j=1}^K \hat\lambda_{nKj} \le C_1\mathbb{P}_n(C_2 + \sum_{j=1}^K y_{Kj})^2 + C_3$, where the right-hand side converges to a finite number by Condition A3 and the strong law of large numbers. On the other hand, we have

$$\begin{aligned}
\limsup_{n\to\infty} \mathbb{P}_n\sum_{j=1}^K \hat\lambda_{nKj} &= \limsup_{n\to\infty} \mathbb{P}_n\hat\Lambda_n(T_{K,K}) \\
&\ge \limsup_{n\to\infty} \mathbb{P}_n\{1_{[t,T]}(T_{K,K})\hat\Lambda_n(T_{K,K})\} \\
&\ge \limsup_{n\to\infty} \hat\Lambda_n(t)\mathbb{P}_n\{1_{[t,T]}(T_{K,K})\} \\
&\ge \gamma([t,T])\limsup_{n\to\infty} \hat\Lambda_n(t).
\end{aligned}$$

Hence $\hat{\Lambda}_n(s)$ is uniformly bounded almost surely for $s \in [0, t]$ if $\gamma([t, T]) > 0$ for some $0 < t < T$ or for $s \in [0, T]$ if $\gamma([T]) > 0$. Then by Helly's selection Theorem and the compactness of $\mathcal{F} \times \mathcal{R}_+^3$, it follows that $\hat{\alpha}_n$ has a subsequence $\hat{\alpha}_{n'}$ converging to $\alpha^+ = (\Lambda^+, r^+, \delta^+, \theta^+)$, where Λ^+ is an increasing bounded function defined on $[0, t]$ for a $t < T$ and it can be defined on $[0, T]$ if $\gamma([T]) > 0$. Following the same argument as in proving Theorem 4.1 of [23], we can show that $M(\alpha^+) \geq M(\alpha_0)$. This implies that $\alpha^+ = \alpha_0$ a.e. in μ. Finally, the dominated convergence theorem yields the strong consistency of $\hat{\alpha}_n$ in the metric d. $\square$

Proof of Theorem 3.2. We derive the rate of convergence of $\hat{\alpha}_n = (\hat{\Lambda}_n, \hat{r}_n, \hat{\delta}_n\hat{\theta}_n)$ by checking the conditions in Theorem 3.2.5 or Corollary 3.2.6 of van der [29]. Since α_0 is the maximum of $M(\alpha)$, then the first derivative is zero at α_0 and the second derivative is negative definite. Thus, for α in a neighborhood of α_0, then there exists a constant C such that $M(\alpha) - M(\alpha_0) \leq -Cd^2(\alpha, \alpha_0)$.

Let $M_\rho = \{m_\alpha(X) - m_{\alpha_0}(X) : d(\alpha, \alpha_0) < \rho\}$ be a class of functions. To find the convergence rate, we need to find $\phi(\rho)$ such that $E\sup_{d(\alpha,\alpha_0)<\rho} \|\mathbb{G}_n\|_{M_\rho} \leq C\phi(\rho)$. We shall find the bracket entropy number for class M_δ. Let $\mathcal{F}_\rho = \{\Lambda \in \mathcal{F} : \|\Lambda - \Lambda_0\|_\mu \leq \rho\}$. Since $\mathcal{F}_\rho$ is the class of monotone function, it is well known that the set of all monotone functions possess a bracketing entropy of the order $1/\epsilon$. Therefore, for any $\epsilon > 0$, there exists a set of brackets $[\Lambda_1^l, \Lambda_1^u], \ldots, [\Lambda_q^l, \Lambda_q^u]$ with $q < \exp(M_1/\epsilon)$, such that for any $\Lambda \in \mathcal{F}_\rho$, $\Lambda_i^l(t) < \Lambda(t) < \Lambda_i^u(t)$ for all $t \in [T_l, T_u]$ for some i and $\|\Lambda_i^u - \Lambda_i^l\|_\mu^2 \leq \epsilon^2$. From Lemma 8.2 in [24], we also can make these bracketing functions satisfying that $\Lambda_i^u - \Lambda_i^l \leq \gamma_1 = 2\epsilon_2$ and $\Lambda_i^l \geq \gamma_2 = \Lambda_0(T_l) - \epsilon_2$ with $\epsilon_2 = (\sqrt{\epsilon^2 + \delta^2}/C)^{2/3}$ for all $t \in [T_l, T_u]$ and i for sufficient small ϵ and ρ. Moreover, by Conditions B3 and B4, there exists a constant $a > 0$ such that $\Lambda_0(T_{K,j}) - \Lambda_0(T_{K,j-1}) > 2a$, and hence we have $\Lambda_i^l(T_{K,j}) - \Lambda_i^u(T_{K,j-1}) \geq \Lambda_0(T_{K,j}) - \Lambda_0(T_{K,j-1}) - 2\epsilon_2 \geq 2(a - \epsilon_2) \geq \gamma_3$ for all $i = 1, \ldots, q$ and $j = 1, \ldots, K$.

Since r is in a compact set, we can construct an ϵ-net for r, $r_1, \ldots, r_p$, with $p = M_2/\epsilon$ such that for any r there is s such that $|r_s - r| \leq \epsilon$. Similarly we have an ϵ-net for $\delta, \delta_1, \ldots, \delta_p$ and an ϵ-net for $\theta, \theta_1, \ldots, \theta_p$. We can construct a set of brackets for M_ρ: $[m_{i,s}^l, m_{i,s}^u]$, $i = 1, \ldots, q$ and $s = 1, \ldots, p$, where

$$m_{i,s}^l = \log\frac{\Gamma(\delta_s - \epsilon + \frac{K}{2})}{\Gamma(\delta_s + \epsilon)} - \frac{K}{2}\log r_s + \epsilon - \frac{1}{2}\log\left(1 + (\theta_s + \epsilon)\sum_{j=1}^K \lambda_i^u\right) - \frac{1}{2}\sum_{j=1}^K \log\lambda_{j,i}^u$$
$$- \left(\delta_s + \epsilon + \frac{K}{2}\right)\log\left(1 + \frac{1}{2(r_s - \epsilon)}\left[-\frac{\left(\sum_{j=1}^K y_j + (\theta_s + \epsilon)^{-1}\right)^2}{\sum_{j=1}^K \lambda_{j,i}^u + (\theta_s - \epsilon)^{-1}} + \sum_{j=1}^K \frac{y_j^2}{\lambda_{j,i}^l} + (\theta_s + \epsilon)^{-1}\right]\right) - m_{\alpha_0}(X)$$

and

$$m_{i,s}^u = \log\frac{\Gamma(\delta_s + \epsilon + \frac{K}{2})}{\Gamma(\delta_s - \epsilon)} - \frac{K}{2}\log r_s - \epsilon - \frac{1}{2}\log\left(1 + (\theta_s - \epsilon)\sum_{j=1}^K \lambda_{j,i}^l\right) - \frac{1}{2}\sum_{j=1}^K \log\lambda_{j,i}^l$$
$$- \left(\delta_s - \epsilon + \frac{K}{2}\right)\log\left(1 + \frac{1}{2(r_s + \epsilon)}\left[-\frac{\left(\sum_{j=1}^K y_j + (\theta_s - \epsilon)^{-1}\right)^2}{\sum_{j=1}^K \lambda_{j,i}^l + (\theta_s + \epsilon)^{-1}} + \sum_{j=1}^K \frac{y_j^2}{\lambda_{j,i}^u} + (\theta_s - \epsilon)^{-1}\right]\right) - m_{\alpha_0}(X),$$

where $\lambda_{j,i}^l = \Lambda_i^l(T_{K,j}) - \Lambda_i^u(T_{K,j-1})$ and $\lambda_{j,i}^u = \Lambda_i^u(T_{K,j}) - \Lambda_i^l(T_{K,j-1})$. In the following, we show that $\|m_{i,j,s}^u - m_{i,j,s}^l\|_{P,B} \leq C\epsilon^2$ where $\|\cdot\|_{P,B}$ is the "Bernstein norm" defined by

$$\|f\|_{P,B} = \sqrt{2P(e^{|f|} - 1 - |f|)}.$$

Since $2(e^x - 1 - x) \leq x^2e^x$ for $x > 0$, we have $\|f\|_{P,B}^2 \leq P(e^{|f|}|f|^2)$. With simple algebra, we can see that $m_{i,j,s}^u - m_{i,j,s}^l$ are all uniformly bounded and there exists a constant C such that

$$\|m_{i,s}^u - m_{i,s}^l\|_{P,B} \leq C\epsilon^2.$$

This shows that the total number of ϵ-brackets for M_ρ will be of order $M_1/\epsilon\exp(CM_2/\epsilon)$ and

$$\log N_{[\,]}(\epsilon, \tilde{M}_\rho, \|\cdot\|_{P,B}) \leq \frac{C}{\epsilon}.$$

Similarly, we can show that $P(m_\alpha(X) - m_{\alpha_0}(X)) \le C\rho^2$ for any $m_\alpha(X) - m_{\alpha_0}(X) \in M_\rho(\alpha_0)$. By Lemma 3.4.3 of van der [29] or Lemma 8.3 of van der [30],

$$E_P^*\|\mathbb{G}_n\|_{M_\rho} \le CJ_{[\,]}(\rho, M_\rho, \|\cdot\|_{P,B})\left(1 + \frac{J_{[\,]}(\rho, M_\rho, \|\cdot\|_{P,B})}{\rho^2\sqrt{n}}\right),$$

where

$$J_{[\,]}(\rho, M_\rho, \|\cdot\|_{P,B}) = \int_0^\rho \sqrt{1 + \log N_{[\,]}(\epsilon, M_\rho(\alpha_0), \|\cdot\|_{P,B})}\mathrm{d}\epsilon$$
$$= C\int_0^\rho \sqrt{1 + \frac{1}{\epsilon}}\mathrm{d}\epsilon \le C\int_0^\rho \epsilon^{-1/2}\mathrm{d}\epsilon \le C\rho^{1/2}.$$

So, $\phi_n(\rho) = \rho^{1/2}(1+\rho^{1/2}/(\rho^2\sqrt{n})) = \rho^{1/2}+\rho^{-1}/\sqrt{n}$, and $\phi_n(\rho)/\rho$ is a decreasing function of ρ, and $n^{2/3}\phi_n(n^{-1/3}) = 2n^{1/2}$. So, by Theorem 3.2.5 of van der [29], we have $n^{1/3}d(\hat{\alpha}_n, \alpha_0) = O_p(1)$. □

Proof of Theorem 3.3. The consistency and the convergence rate of the bootstrap estimators can be shown by imitating the proofs in Theorems 3.1 and 3.2. We omit the details here. □

References

[1] V. DeGruttola, N. Lange, U. Dafni, Modeling the progression of HIV infection, Journal of the American Statistical Association 86 (1991) 569–577.
[2] S.J. Hudak Jr., A. Saxena, R.J. Bucci, R.C. Malcolm, Development of standard methods of testing and analyzing fatigue crack growth rate data, Technical Report AFML-TR-78-40, Westinghouse R& D Center, Westinghouse Electric Corporation, PA 15235, Pottsburgh, 1978.
[3] C.J. Lu, W.Q. Meeker, Using degradation measures to estimate a time-to-failure distribution, Technometrics 35 (1993) 161–171.
[4] E.A. Elsayed, H.T. Liao, A geometric Brownian motion model for field degradation data, International Journal Materials and Product Technology 20 (2004) 51–72.
[5] N.D. Singpurwalla, Survival in dynamic environments, Statistical Science 1 (1995) 86–103.
[6] W.Q. Meeker, L.A. Escobar, Statistical Methods for Reliability Data, Wiley, New York, 1998.
[7] X. Wang, Physical degradation models, in: F. Ruggeri, R. Kenett, F.W. Faltin (Eds.), Encyclopedia of Statistics in Quality and Reliability, John Wiley & Sons Lts, Chichester, UK, 2007, pp. 1356–1361.
[8] K. Doksum, Degradation rate models for failure time and survival data, CWI Quarterly (1991) 195–203.
[9] K. Doksum, A. Hoyland, Models for variable stress accelerated life testing experiments based on wiener process and the inverse gaussian distribution, Technometrics 34 (1992) 74–82.
[10] G.A. Whitmore, Estimation degradation by a Wiener diffusion process subject to measurement error, Lifetime Data Analysis 1 (1995) 307–319.
[11] G.A. Whitmore, F. Schenkelberg, Modelling accelerated degradation data using Wiener diffusion with a time scale transformation, Lifetime Data Analysis 3 (1997) 27–43.
[12] R.S. Chikkara, J.L. Folks, The Inverse Gaussian Distribution, Marcell Dekker, NY, 1989.
[13] V. Sheshadri, The Inverse Gaussian Distribution, Springer, NY, 1998.
[14] X. Wang, Nonparametric inference with applications to dark matter estimation in astronomy and degradation modeling in reliability, Ph.D. dissertation, University of Michigan, 2005.
[15] K. Chaloner, I. Verdinelli, Bayesian experimental design: A review, Statistical Science 10 (1995) 273–304.
[16] J. Lawless, M. Crowder, Covariate and random effects in a Gamma process model with application to degradation and failure, Lifetime Data Analysis 10 (2004) 213–227.
[17] R.J.A. Little, D.B. Rubin, Statistical Analysis with Missing Data, John Wiley, New York, 1987.
[18] R.J.A. Little, Robust estimation of the mean and covariance matrix from data with missing values, Applied Statistics 37 (1988) 23–38.
[19] K.L. Lange, R.J.A. Little, J.M.G. Taylor, Robust statistical modeling using the t distribution, Journal of the American Statistical Association 84 (1989) 881–896.
[20] C. Liu, D.B. Rubin, ML estimation of the multivariate t distribution with unknown degrees of freedom, Statistica Sinica 5 (1995) 19–39.
[21] C. Liu, ML estimation of the multivariate t distribution and the EM algorithm, Journal of Multivariate Analysis 63 (1997) 296–312.
[22] D.P. Bertsekas, Nonlinear Programming, 2nd ed., Athena Scientific, 2003.
[23] J. Wellner, Y. Zhang, Two estimators of the mean of a counting process with panel count, The Annals of Statistics 28 (2000) 779–814.
[24] J. Wellner, Y. Zhang, Two likelihood-based semiparametric estimation methods for panel data with covariates, The Annals of Statistics 35 (2007) 2106–2142.
[25] B. Efron, R. Tibshirani, An Introduction to Bootstrap, Chapman & Hall/CRC, 1993.
[26] B. Efron, Nonparametric standard errors and confidence intervals (with discussion), Canadian Journal of Statistics 9 (1981) 139–172.
[27] B. Efron, The Jackknife, the Bootstrap, and Other Resampling Plans, SIAM, Philadelphia, 1982.
[28] V. Bagdonavicius, M. Nikulin, Estimation in degradation models with explanatory variables, Lifetime Data Analysis 7 (2000) 85–103.
[29] van der Vaart, J. Wellner, Weak Convergence and Empirical Processes with Applications to Statistics, in: Springer Series in Statistics, 1996.
[30] van der Vaart, Semiparametric statistics, in: Lectures on Probability Theory and Statistics, in: Lecture Notes in Mathematics, vol. 1781, 2002, pp. 330–457.
[31] B. Jorgensen, Statistical Properties of the Generalized Inverse Gaussian Distribution, Springer-Verlag, New York, 1982.