A General Class of Models for Recurrent Events

Edsel A. Pena

University of South Carolina at Columbia

[E-Mail: pena@stat.sc.edu]

SRC, U of M, Ann Arbor, MI, 5/21/02

Research support from NIH, NSF

Recurrent Phenomena

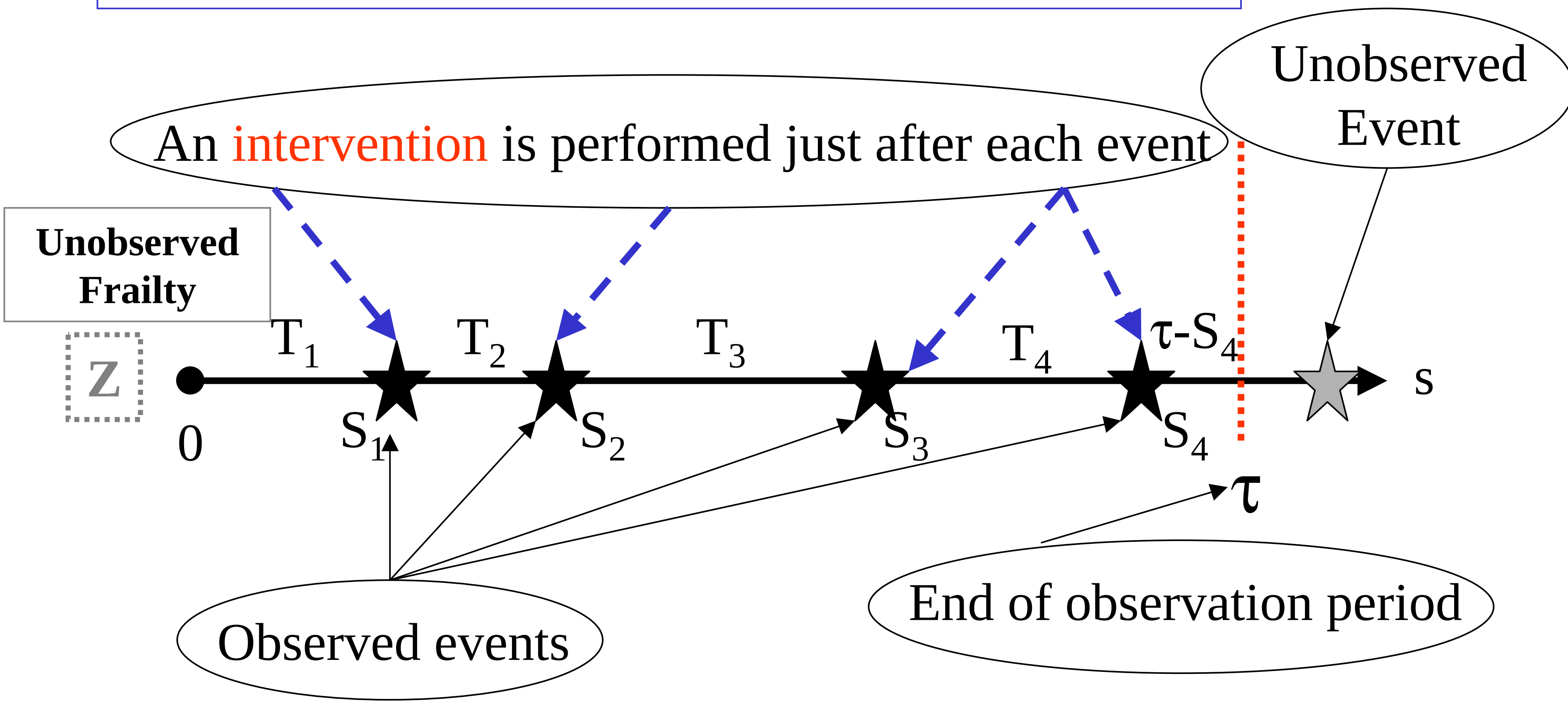
In Reliability, Engineering, and Economic Settings

- failure of a mechanical/electronic system
- warranty claims
- Dow Jones index changes by more than 200 points
- occurrence of a certain type of accident (nuclear)
- occurrence of a terrorist attack

In Public Health and Medical Settings

- hospitalization of a subject with a chronic disease
- tumor occurrence
- cyclic movements in the small bowel during fasting state
- episodes of depression

Data Accrual for a Subject/Unit



An observable covariate vector: $\mathbf{X}(s) = (X_1(s), X_2(s), \ldots, X_q(s))^t$

Features in Recurrent Event Modeling

- Intervention (repair) effects after each event occurrence.
- Effects of accumulating event occurrences on the subject. Could be a weakening or an strengthening effect.
- Effects of possibly time-dependent covariates.
- Possible associations of event occurrences for a subject.
- A possibly random observation period per subject.
- Number of observable events per subject is random and is informative on stochastic mechanism generating events.
- Informative right-censoring mechanism for the inter-event time that covers end of observation period.

Random Entities: One Subject

- T_1, T_2, T_3, … = the inter-event or gap times
- S_1, S_2, S_3, … = calendar times of occurrences
- $\mathbf{X}(s)$ = covariate vector, possibly time-dependent
- $\mathbf{F}^+ = \{F_s^+: 0 \leq s\}$ = filtration including info about interventions, covariate, etc. in [0, s]
- Z = unobserved frailty (latent) variable
- $N^+(s)$ = number of events observed on or before calendar time s
- $Y^+(s)$ = indicator of whether the subject is still at risk just before calendar time s

A General Class of Models

$\{A^+(s|Z): s \geq 0\}$ is a predictable non-decreasing process such that, given Z, and with respect to the filtration $\mathbf{F}^+$:

$$\left\{M^+(s\,|\,Z) = N^+(s) - A^+(s\,|\,Z):\ s \geq 0\right\}$$

is a square-integrable zero-mean (local) martingale. As in previous works (Aalen, Gill, Andersen and Gill, Cox, Nielsen, et al, others) we assume

$$A^+(s\,|\,Z) = \int_0^s Y^+(w)\lambda(w\,|\,Z)\,dw$$

Modeling the Intensity Process [Pena and Hollander, to appear]

Specify, possibly in a dynamic fashion, a predictable, observable process $\{E(s): 0 \leq s \leq \tau\}$, called the *effective age process*, satisfying

- $E(0) = e_0 \geq 0$;
- $E(s) \geq 0$ for every s;
- On $[S_{k-1}, S_k)$, $E(s)$ is monotone and differentiable

with a nonnegative derivative.

Specification of the Intensity Process

$$\lambda(s \mid Z) = Z\lambda_0[E(s)]\rho[N^+(s-);\alpha]\psi[\beta^t X(s)]$$

Frailty

Baseline Hazard

Effective Age, encodes intervention effects

Encodes Effect of Accumulating Events

Encodes effects of Covariates or group memberships

Model Components

• $\lambda_0(.)$ = an unknown baseline hazard rate function, possibly parametrically specified.

• E(s) = ***effective age*** of the subject at calendar time s. Idea is that a performed intervention changes the effective age of subject acting on the baseline hazard rate.

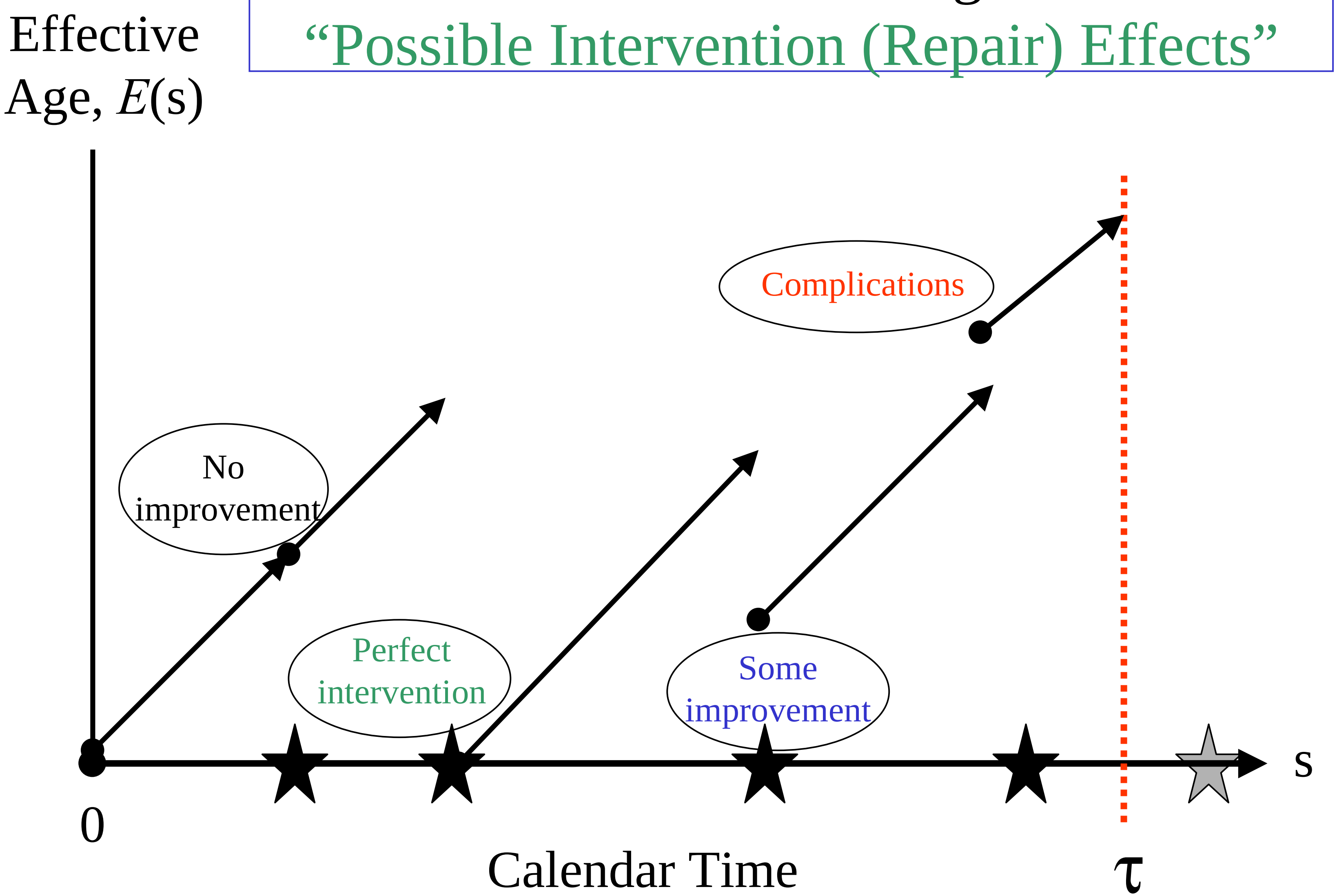
• $\rho(.;\alpha)$ = a +function on {0,1,2,…} of known form with $\rho(0;\alpha) = 1$ and with unknown parameter α. Encodes effect of accumulating event occurrences on the subject.

• $\psi(.)$ = positive link function containing the effect of subject covariates. β is unknown.

• Z = unobservable frailty variable, which when integrated out, induces associations among the inter-event times.

Illustration: Effective Age Process

"Possible Intervention (Repair) Effects"



Special Cases of the Class of Models

• Renewal (IID) Model <u>without</u> frailties: Considered by Gill ('81 AS), Wang and Chang ('99, JASA), Pena, Strawderman and Hollander ('01, JASA).

$$E(s) = s - S_{N^{+}(s-)}; Z = 1; \rho(k;\alpha) = 1; \psi(w) = 1.$$

• Renewal (IID) Model <u>with</u> frailties: Considered by Wang and Chang ('99), PSH ('01).

$$E(s) = s - S_{N^{+}(s-)}; Z \sim Ga(\gamma,\gamma); \rho(k;\alpha) = 1; \psi(w) = 1.$$

Generality and Flexibility

- Extended Cox PH Model: Considered by Prentice, Williams, and Petersen (PWP) (‘81); Lawless (‘87), Aalen and Husebye (‘91).

$$E(s) = s - S_{N^{+}(s-)}; Z = 1; \rho(k;\alpha) = 1; \psi(w) = exp(w).$$

- Also by PWP (‘81), Brown and Proschan (‘83) and Lawless (‘87) called a “minimal repair model” in the reliability literature.

$$E(s) = s; Z = 1; \rho(k;\alpha) = 1; \psi(w) = 1.$$

A Tumor Occurrence Model and a Software Reliability Model

- A generalized Gail, Santner and Brown ('80) tumor occurrence model;
- Jelinski and Moranda ('72) software reliability model:

$$\mathrm{E(s)} = \mathrm{s} - \mathrm{S}_{\mathrm{N}^{+}(s-)};$$

$$\rho(k;\alpha) = \alpha - k + 1;$$

$$Z = 1;$$

$$\psi(\mathrm{w}) = \exp(w).$$

Generalized Minimal Repair Models

• Let I_1, I_2, I_3, … be independent Ber[p(s)] rvs and $\eta(s) = \sum_{i=1}^{N^+(s)} I_i$. Let $\Gamma_k = \min\{j > \Gamma_{k-1}: I_j = 1\}$. If

$$E(s) = s - S_{\Gamma_{\eta(s-)}}$$

the BP (‘83) and Block, Borges and Savits (‘85) minimal repair model obtains. Also considered in Presnell, Hollander and Sethuraman (‘94, ‘97) and Whitaker and Samaniego (‘89).

Other Models In Class

• Dorado, Hollander and Sethuraman (‘97), Kijima (‘89), Baxter, Kijima and Tortorella (‘96), Stadje and Zuckerman (‘91), and Last and Szekli (‘98):

$$\{A_j : j = 0,1,2,\ldots\} \text{ and } \{\Theta_j : j = 0,1,2,\ldots\}$$

$$E(s) = A_{N^+(s-)} + \Theta_{N^+(s-)}\left[s - S_{N^+(s-)}\right]$$

Forms of ρ

Two simple forms for the ρ function:

$$\rho(k;\alpha) = \alpha^k;$$
$$\rho(k;\alpha) = \max\{\alpha - g(k), 0\}$$

α = initial measure of “defectiveness” or event “proneness.”

Relevance

- ***Flexibility*** and ***generality*** of class of models will allow better modeling of observed phenomena, and allow testing of specific/special models using this ***larger*** class.
- **Question: Is this relevant in reliability, engineering, or biostatistical modeling??**
- **Answer:** The ***fact*** that it contains models currently being used indicates the model's importance.
- **However**, a ***"paradigm shift"*** is needed in the data gathering since the model requires the assessment of the effective age.
- **But**, this could be provided by the reliability, engineering, and medical/public health experts after each intervention.

On the <u>General</u> Class of Model's *Immediate* Applicability

Most often it is the case of

"A Data in Search of a Model;"

but, sometimes* as in this case, it is

"A Model in Search of a Data!"

*A modern example of such a situation is that which led to the 1919 Eddington expedition.

A *Crucial* Inference Issue

- Must take into account the sum-quota data accrual scheme, which leads to an:

- informative random number of events;
- informative right-censoring mechanism.

- Related to the issue of selection bias.

Special Case: Renewal (IID) Model

(Case with: $E(s) = s - S_{N^+(s-)}$; $Z = 1$; $\rho = 1$; $\psi = 1$)

Unit #	Successive Inter-Event Times or Gaptimes	Length of Study Period
1	$T_{11}, T_{12}, \ldots, T_{1j}, \ldots$ IID F	τ_1
2	$T_{21}, T_{22}, \ldots, T_{2j}, \ldots$ IID F	τ_2
…	…	…
n	$T_{n1}, T_{n2}, \ldots, T_{nj}, \ldots$ IID F	τ_n

Calendar Time of Occurrences: $S_{ij} = T_{i1} + T_{i2} + \ldots + T_{ij}$

Number of Events in Obs. Period: $K_i = \max\{j: S_{ij} \le \tau_i\}$

G = common distribution function of the study period lengths

MMC Data: A Real Recurrent Event Data

(Source: Aalen and Husebye ('91), *Statistics in Medicine*)

Variable: Migrating motor complex (MMC) periods, in minutes, for 19 individuals in a gastroenterology study concerning small bowel motility during fasting state.

Unit # i	#Complete ($K_i = K(i)$)	Complete Observed Successive Periods (T_{ij})	Censored ($\tau_i - S_{iK(i)}$)
1	8	112 145 39 52 21 34 33 51	54
2	2	206 147	30
3	3	284 59 186	4
4	3	94 98 84	87
5	1	67	131
6	9	124 34 87 75 43 38 58 142 75	23
7	5	116 71 83 68 125	111
8	4	111 59 47 95	110
9	4	98 161 154 55	44
10	2	166 56	122
11	5	63 90 63 103 51	85
12	4	47 86 68 144	72
13	3	120 106 176	6
14	4	112 25 57 166	85
15	3	132 267 89	86
16	5	120 47 165 64 113	12
17	4	162 141 107 69	39
18	6	106 56 158 41 41 168	13
19	5	147 134 78 66 100	4

Renewal Model Setting: Notations

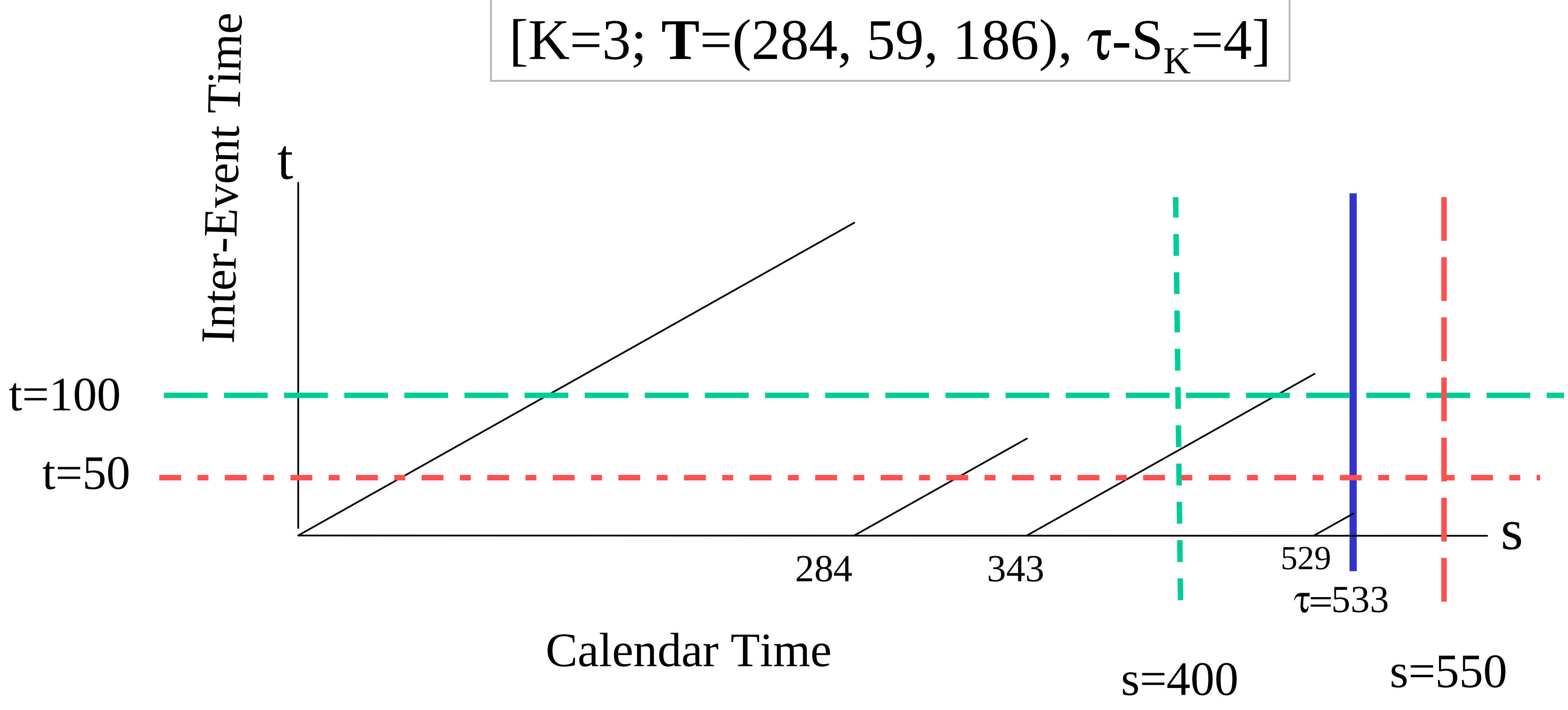
$N_i(s,t)$ = number of events for the i^{th} unit in calendar period [0,s] with inter-event times at most t.

$Y_i(s,t)$ = number of events for the i^{th} unit which are known during the calendar period [0,s] to have inter-event times at least t.

$K_i(s)$ = number of events for the i^{th} unit that occurred in [0,s].

For MMC Unit #3

[K=3; **T**=(284, 59, 186), τ-S_K=4]



K_3(s=400) = 2; N_3(s=400,t=100) = 1; Y_3(s=400,t=100) = 1

K_3(s=550) = 3; N_3(s=550,t=50) = 0; Y_3(s=550,t=50) = 3

Estimating F: Renewal (IID) Model

Aggregated Processes:

$$N(s,t) = \sum_{i=1}^{n} N_i(s,t) \quad \text{and} \quad Y(s,t) = \sum_{i=1}^{n} Y_i(s,t)$$

Limit Processes as s Increases:

$$K_i = K_i(\infty)$$

$$N_i(t) = N_i(\infty, t) = \sum_{j=1}^{K_i} I\{T_{ij} \le t\}$$

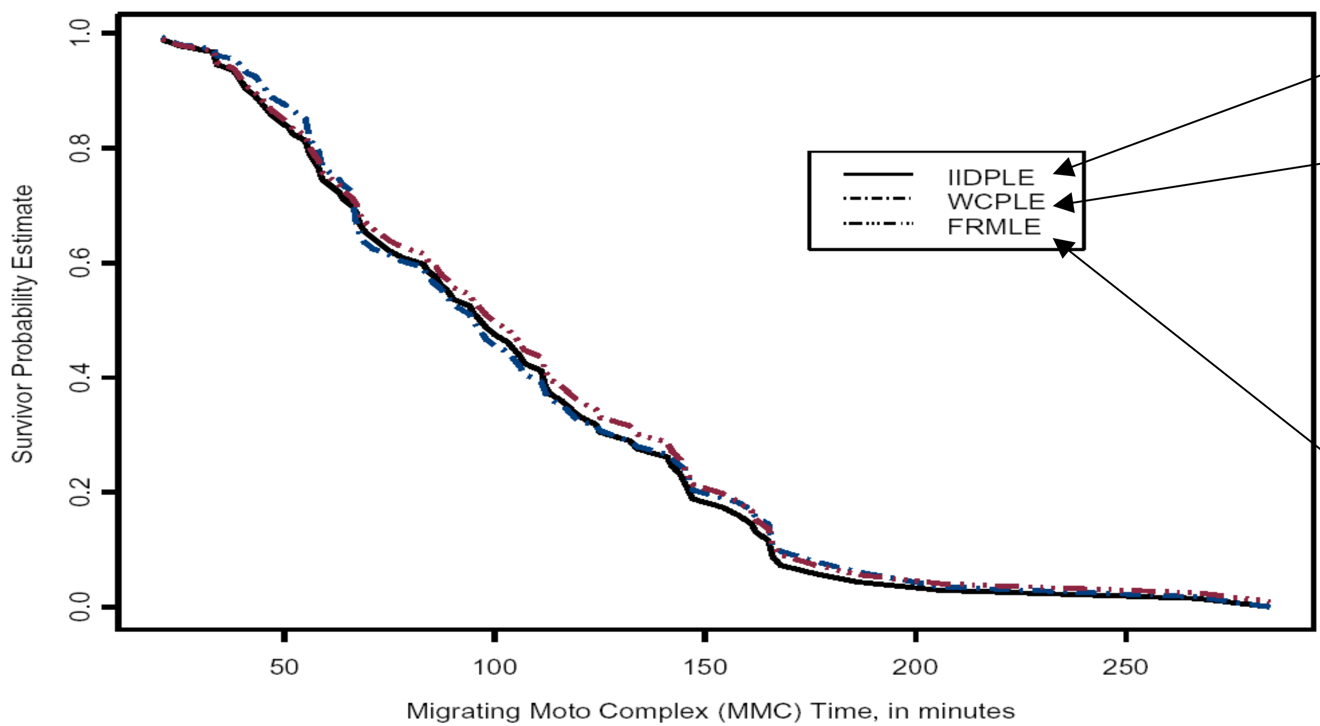
$$Y_i(t) = Y_i(\infty, t) = \sum_{j=1}^{K_i} I\{T_{ij} \ge t\} + I\{\tau_i - S_{iK(i)} \ge t\}$$

Generalized PLE [PSH, *JASA* (‘01)] in Renewal Model (for s large)

$$\hat{\bar{F}}(t) = \prod_{\{w: w \le t\}} \left[1 - \frac{\Delta N(w)}{Y(w)} \right] = \prod_{i=1}^{n} \prod_{\{j: T_{ij} \le t\}} \left[1 - \frac{1}{\sum_{l=1}^{n} Y_l(T_{ij})} \right]$$

Estimator is called the GPLE or the IIDPLE; generalizes the empirical survivor function (EDF) and the product-limit estimator (PLE).

Illustration: Three Estimates of the MMC Period Survivor Function



GPLE

Wang-Chang estimator (JASA, '99)

Gamma frailty-based estimator

Asymptotic Properties of GPLE

$$F^{*j} = j\text{th convolution of F} = \text{dist of } (T_1 + T_2 + \ldots + T_j)$$

$$R(t) = \text{renewal function of F} = \sum_{\text{j=1}}^{\infty} F^{*j}(t)$$

$$y(t) = \overline{F}(t)\overline{G}(t)\left[1 + \int_t^{\infty} R(w - t)\,dG(w \mid \tau \geq t)\right]$$

As n increases: $\sqrt{n}\left(\hat{\overline{F}}(t) - \overline{F}(t)\right) \sim N(0, \upsilon_3(t))$

$$\upsilon_3(t) = \overline{F}(t)^2 \int_0^t \frac{d\Lambda(w)}{y(w)}; \qquad \Lambda = -\log(1 - F)$$

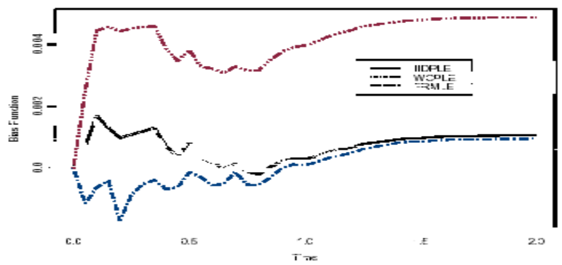
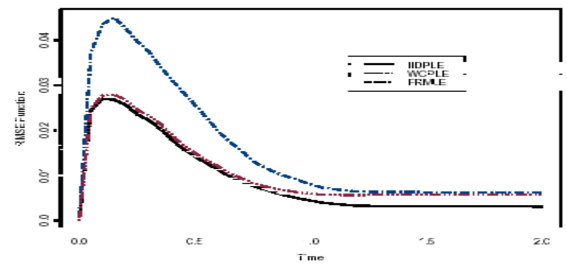
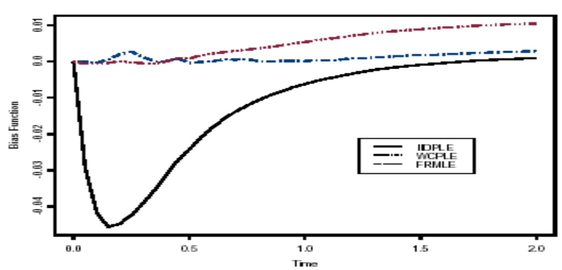
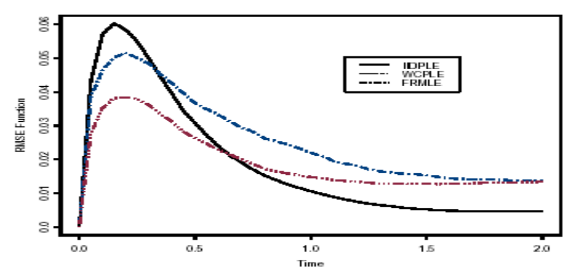
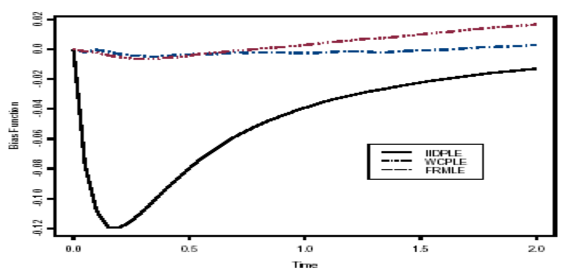
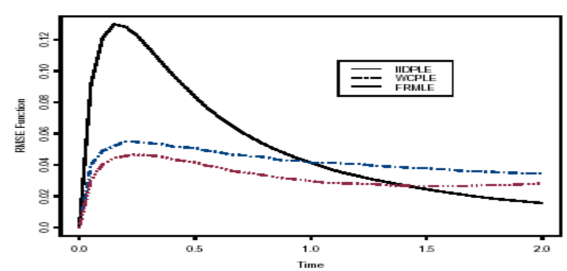
Evolution: Limiting Variances

$$\text{EDF}: \quad v_1(t) = F(t)\overline{F}(t) = \overline{F}(t)^2 \int_0^t \frac{d\Lambda(w)}{\overline{F}(w)}$$

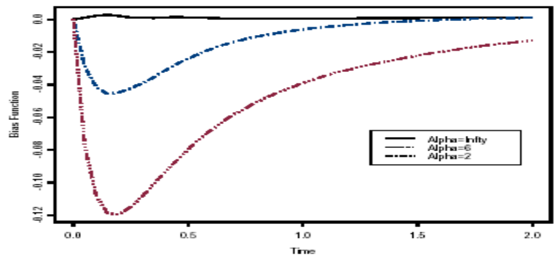
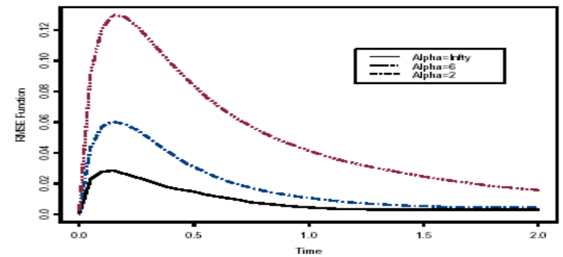
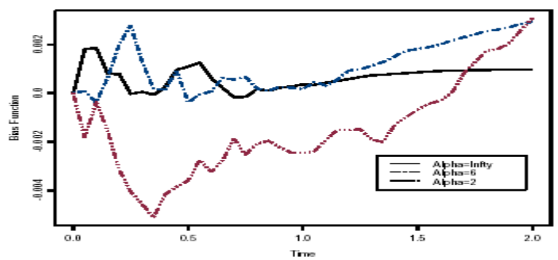
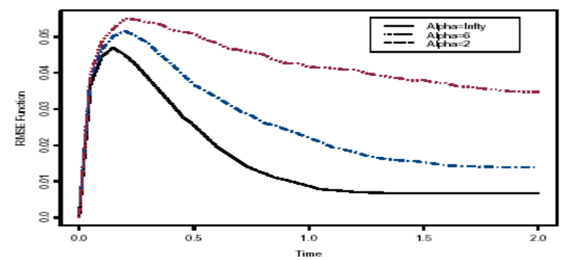
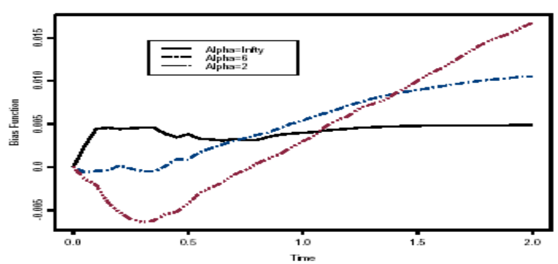
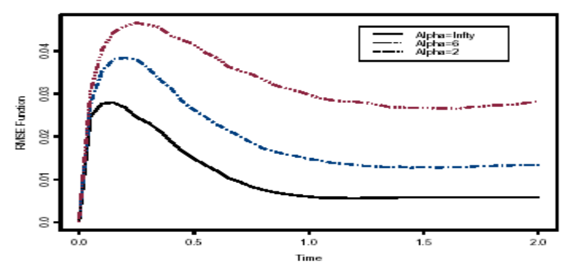
$$\text{PLE}: \quad v_2(t) = \overline{F}(t)^2 \int_0^t \frac{d\Lambda(w)}{\overline{F}(w)\overline{G}(w)}$$

$$\text{GPLE:} \quad v_3(t) = \bar{F}(t)^2 \int_0^t \frac{\mathrm{d}\Lambda(w)}{\bar{F}(w)\bar{G}(w)\left\{1 + \frac{1}{\bar{G}(w)} \int_w^\infty R(u-w)\mathrm{d}G(u)\right\}}$$

Comparison of Three Estimators: Varying Frailty Parameter
Black=GPLE; Blue=WCPLE; Red=FRMLE

Frailty Parameter α	Simulated Bias Function	Simulated RMSE Function
∞ (Indep)		
6 (Moderate)		
2 (Strong)		

As the Frailty Parameter (α) or Association Changes (Black=Indep.; Blue=Moderate; Red=Strong)

Estimator	Simulated Bias Function	Simulated RMSE Function
IIDPLE		
WCPLE		
FRMLE		

Inference Problems: General Model

- Parameter Estimation.
- Testing and Group Comparisons.
- Testing the Frailty Model Assumption.
- Model Validation and Diagnostics.
- Application to real data where the effective age process is monitored.