Change-Point Analysis of Survival Data with application in Clinical Trials

by

Xuan Chen, Michael Baron

Department of Mathematical Sciences
University of Texas at Dallas
Richardson, Texas 75083-0688

June 04, 2009

Outline

- **Introduction to classical Change-Point Problem**
- **Change-Point Analysis of Survival Data**
- **Constant Hazard Rates Model with a change-point**
  - Maximum Likelihood Method
  - Least Squares Method
  - Comparison through a Simulation Study
- **Application in a recent Clinical Trial**

Classical Change-Point Problem

- Observe

$$\boldsymbol{X} = (X_1, \ldots, X_\tau, X_{\tau+1}, \ldots, X_n) = (\boldsymbol{X}_1, \boldsymbol{X}_2)$$

$$\begin{array}{lll} X_1, \ldots, X_\tau & \text{from} & f(x) \\ X_{\tau+1}, \ldots, X_n & \text{from} & g(x) \end{array}$$

  where τ = change-point parameter

- Application
  - quality control / epidemiology / financial data
- Theory : MLE, LRT, CUSUM, Bayesian, Exponentially Weighted Moving Average
- Maximum Likelihood Estimate
  - MLE of τ is not consistent
  - MLE of τ is asymptotically the uniform minimum variance unbiased estimator

Survival Data with a Change-Point

- Piece-wise Constant Hazard Model

$$\lambda(x) = \lambda_0 1_{x \le \tau} + \lambda_1 1_{x > \tau} \tag{1}$$

- Survival Times: $X_1, \ldots, X_n$ iid with pdf $f(x)$
- Censoring Time: t
- Joint Density Function

$$\begin{aligned}
& f(x_1, ..., x_n | \lambda_0, \lambda_1, \tau) \\
& = \lambda(x) \text{exp} \left\{ - \int_0^x \lambda(t) dt \right\} \\
& = \prod_{x_i \le t} (\lambda_0 e^{-\lambda_0 x_i} 1_{x_i \le \tau} + \lambda_1 e^{-\lambda_0 \tau - \lambda_1 (x_i - \tau)} 1_{x_i > \tau}) \\
& \quad \prod_{x_i > t} (e^{-\lambda_0 t} 1_{t \le \tau} + e^{-\lambda_0 \tau - \lambda_1 (t - \tau)} 1_{t > \tau})
\end{aligned}$$

Maximum Likelihood Estimate

(Without Nuisance Parameters)

Maximize

$$\begin{aligned}\Lambda_t &= \log \frac{f(x_1, ..., x_n|\lambda_0, \lambda_1, \tau)}{f(x_1, ..., x_n|\lambda_0, \tau = \infty)} \\ &= \sum_{i \le n} y_i\end{aligned}$$

where

$$y_i = \begin{cases} \log\frac{\lambda_1}{\lambda_0} + (\lambda_0 - \lambda_1)(x_i - \tau) & \text{for } \tau < x_i \le t \\ (\lambda_0 - \lambda_1)(t - \tau) & \text{for } \tau < t < x_i \\ 0 & \text{otherwise} \end{cases}$$

Note: Λ_t is piecewise linear in τ and maximum of Λ_t should achieve on some X_i.

$$\hat{\tau} = \{X_{(k)} : \max_k \sum_{i=k+1}^{n} y_{(i)}\}$$

Convergence Rate of $\hat{\tau}$

Theorem 1. *Suppose that the true value of change-point is τ_0. Then the log-likelihood ratio $\Lambda_t(\tau)$ has the following properties*

(i) $E\Lambda_t(\tau)$ *is strictly increasing in τ when $\tau < \tau_0$.*
(ii) $E\Lambda_t(\tau)$ *is strictly decreasing in τ when $\tau_0 < \tau < t$.*
(iii) $\Lambda_t(\tau)$ *is constant after time t.*

Theorem 2. *Maximum likelihood estimator $\hat{\tau}$ converges to τ almost surely. For any $\epsilon > 0$, there exists $N > 0$, such that*

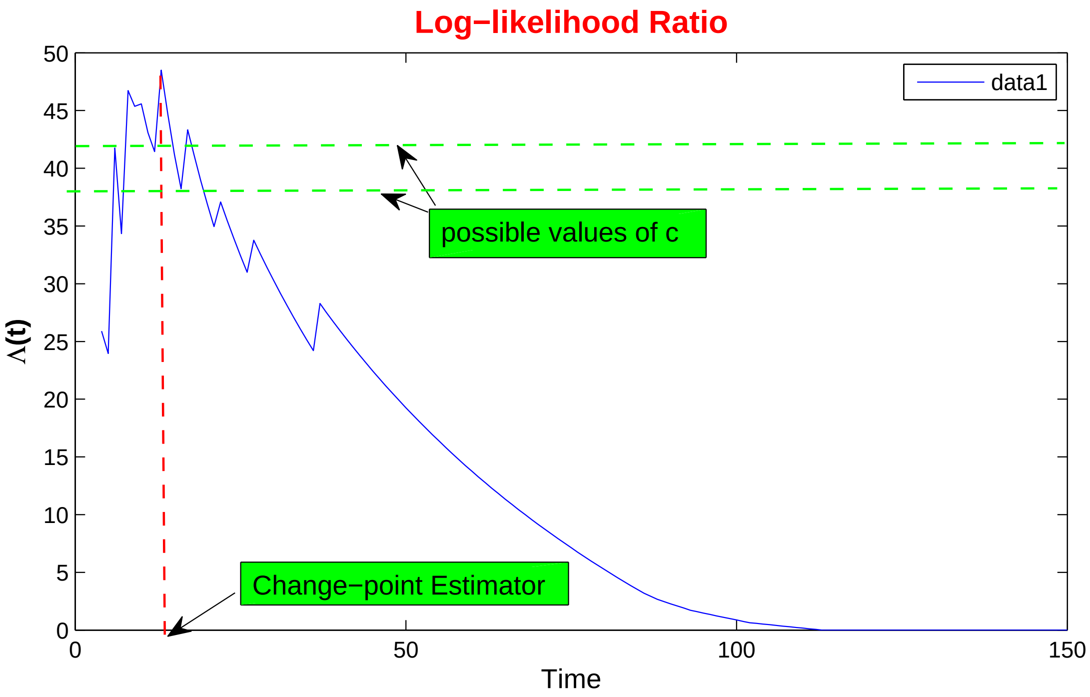
$$\mathbf{P}(\exists\, \tau : |\tau - \tau_0| > \epsilon \text{ and } \Lambda_t(\tau) - \Lambda_t(\tau_0) > 0) \leq \frac{4\sigma_1^N}{1 - \sigma_1},$$

where

$$\sigma_1 = e\rho^{\frac{1}{1-\rho}} \frac{\log \rho}{\rho - 1} < 1,$$

$$\rho = \lambda_1 / \lambda_0.$$

Confidence Region and its Asymptotic Behavior



$R_c = \{\tau : \max_{0 \le s \le t} \Lambda_t(s) - \Lambda_t(\tau) < c\}$ minimizes risk function: $\lambda \mathsf{E}|R_c| + P(\tau \notin R_c)$

Theorem 3. *(Asy. behavior of $P(\tau \in R_c)$)* *For* $c \to \infty$

$$P(\tau \notin R_c) \sim e^{-c}\pi_0\pi_1(\frac{1}{\rho_0} + \frac{1}{\rho_1})$$

where

$$\pi_0 = exp\left\{ - \sum_{k=1}^{\infty} k^{-1} P(S'_k \geq 0) \right\}$$

$$\pi_1 = exp\left\{ - \sum_{k=1}^{\infty} k^{-1} P(S_k > 0) \right\}$$

$$\rho_0 = -E_0 W'_m = (\frac{\lambda_1}{\lambda_0} - 1) - log\frac{\lambda_1}{\lambda_0}$$

$$\rho_1 = -E_1 W_m = (\frac{\lambda_0}{\lambda_1} - 1) - log\frac{\lambda_0}{\lambda_1}. \qquad (2)$$

Corollary 4. *The boundary for level $(1-\alpha)$ confidence region is approximately*

$$c \sim log\frac{\pi_0\pi_1(\frac{1}{\rho_0} + \frac{1}{\rho_1})}{\alpha}$$

Theorem 5. *(Asy. behavior of E$|R_c|$)* *The average length of* R_c *for* $c \to \infty$ *has the form*

$$E|R_c| \sim c[\frac{1}{\rho_0} + \frac{1}{\rho_1}] + b + o(1)$$

where b *is a constant term.*

Theorem 6. *(Asy. behavior of risk)* *As* $c \to \infty$*,*

$$\boldsymbol{E}W(\tau, R_c) = (\lambda c + e^{-c}\pi_0\pi_1)(\frac{1}{\rho_0} + \frac{1}{\rho_1}) + \lambda(b + o(1))$$

Constant c *which minimizes this risk admits the following asymptotic representation for* $\lambda \to 0$*,*

$$c \sim -\log \lambda + \log \pi_0\pi_1.$$

non-symmetric confidence region

$$
\begin{aligned}
R_{c_0,c_1} &= R_{c_0} \cup R_{c_1} \\
&= \{\tau \leq \hat{\tau} : \Lambda_t(\hat{\tau}) - \Lambda_t(\tau) < c_0\} \cup \{\tau > \hat{\tau} : \Lambda_t(\hat{\tau}) - \Lambda_t(\tau) < c_1\}
\end{aligned}
$$

Assume that c_0 and c_1 converge at the same rate. Then as $c_0 \to \infty$, $c_1 \to \infty$

$$
\begin{aligned}
&\boldsymbol{E}W(\tau, R_{c_0,c_1}) \qquad (3) \\
&\sim \lambda\left(\frac{c_0}{\rho_0} + \frac{c_1}{\rho_1} + b + o(1)\right) + \pi_0\pi_1\left(\frac{e^{-c_0}}{\rho_0} + \frac{e^{-c_1}}{\rho_1}\right) \\
&= (\lambda c_0 + e^{-c_0}\pi_0\pi_1)\frac{1}{\rho_0} + (\lambda c_1 + e^{-c_1}\pi_0\pi_1)\frac{1}{\rho_1} + \lambda(b + o(1))
\end{aligned}
$$

Theorem 7. *R_c is optimal among R_{c_0,c_1} in the sense of minimizing asymptotic risk(3).*

Maximum Likelihood Estimate

(With Nuisance Parameters)

- Log-Likelihood function is maximized when

$$\hat{\lambda}_0 = R_\tau(\sum (X_i 1_{X_i \le \tau < t}) + \tau(n - R_\tau))^{-1}$$

$$\hat{\lambda}_1 = R_t - R_\tau(\sum_{X_i \le t} (X_i - \tau)^+ + (t - \tau)(n - R_t))^{-1}$$

- Substitute $\hat{\lambda}_0, \hat{\lambda}_1$ back to $L(\tau) = \log f(\tau, \hat{\lambda}_0, \hat{\lambda}_0)$ and maximize $L(\tau)$ with respect to τ.
- MLE : $\hat{\tau} = \{\tau : \max_\tau L(\tau)\}$.

Theorem 8. $\hat{\tau}, \hat{\lambda}_0, \hat{\lambda}_1$ *are consistent.*

Least Squares Method based on Kaplan-Meier Estimation

- Kaplan and Meier (1958) suggested a nonparametric estimator for the survival function $S(t)$:

$$\tilde{S}_n(x) = \prod_{x_{(j)} \leq x} \left(\frac{n-j}{n-j+1} \right)^{\delta_{(j)}} \tag{4}$$

- Let $\theta = (\lambda_0, \lambda_1, \tau)$ denote the parameter space.
- LSE $\tilde{\tau}, \tilde{\lambda}_0, \tilde{\lambda}_1$ minimize the error sum of squares

$$\text{ESS}(\theta) = \sum_{i=1}^{n} (\tilde{y}_n(x_i) - L_i(\theta))^2, \tag{5}$$

where $\tilde{y}_n(x_i) = \log \tilde{S}_n(x_i)$,
$L_i(\tau, \lambda_0, \lambda_1) = \log S(x_i)$.

Strong Convergency of Least Squares Estimators

Theorem 9. $\tilde{\tau}, \tilde{\lambda}_0, \tilde{\lambda}_1$ *converges almost surely to* $\tau, \lambda_0, \lambda_1$.

Theorem 10. *For any* $\epsilon > 0$, *there exists* $c > 0$, *such that*

$$\mathbf{P}(\exists\, x : |\tau - \tau_0| > \epsilon \text{ and } ESS(\tau) - ESS(\tau_0) < 0) \leq e^{-nc}$$

for sufficiently large n.

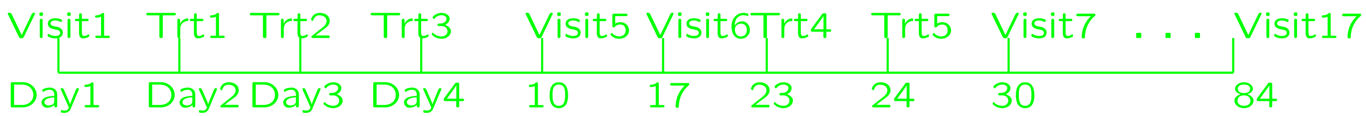
Comparison of MLE and LSE through Simulation Study

- Simulate data from model (1)
- $\tau = 5, t = 20$
  $(\lambda_0, \lambda_1) = (0.2, 0.15), (0.25, 0.15)$, and $(0.3, 0.1)$.
- Samples sizes from 500, 1000, 1500 were considered.
- Conclusions
  - Both MLE and LSE are more accurate when we increase the difference between λ_0 and λ_1 for the same sample size.
  - Both MLE and LSE converge to the true change-point when we increase the sample size for the same λ_0 and λ_1.
  - LSE is more accurate than MLE for the same sample size and same failure rates.

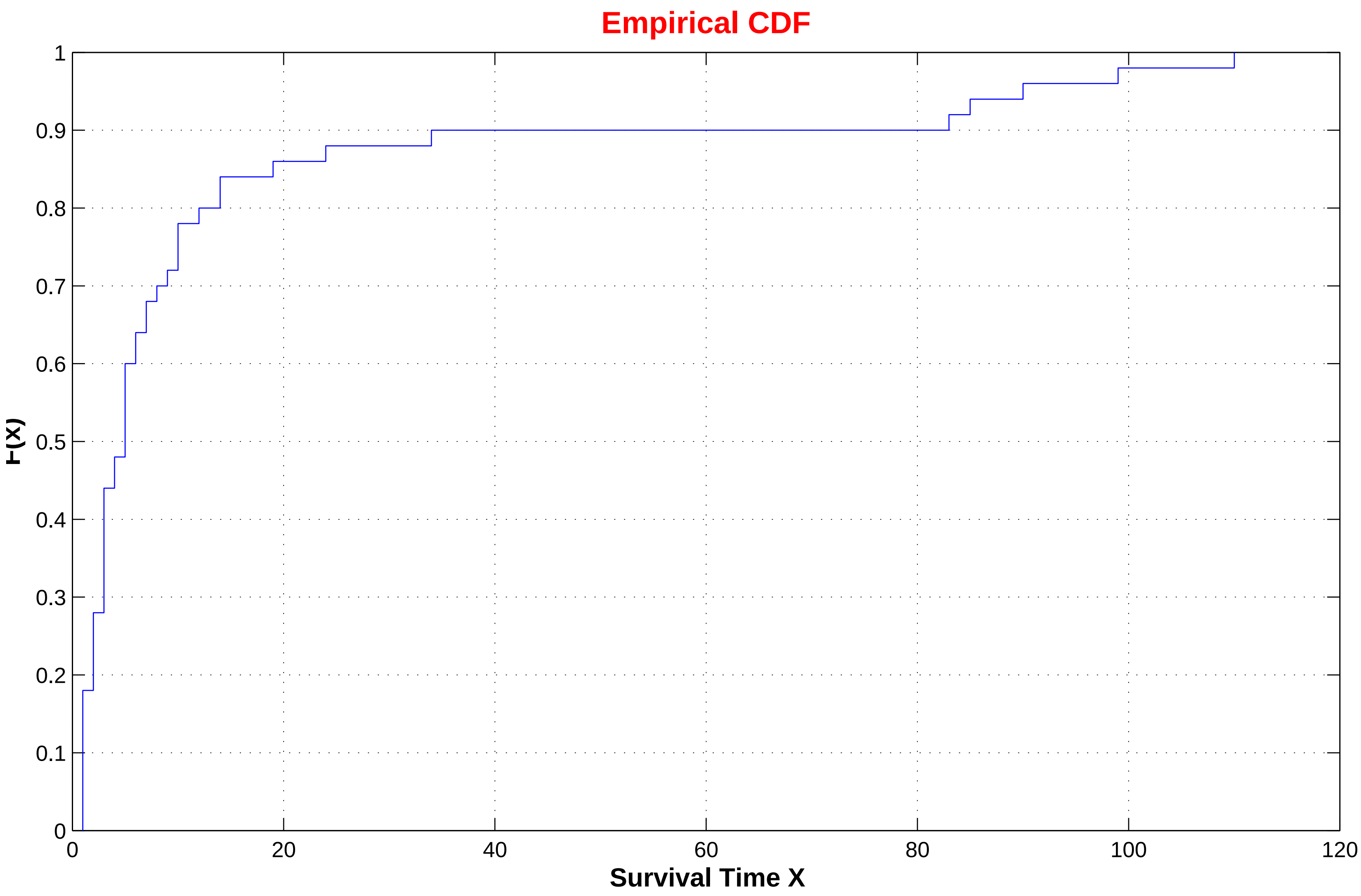
Method	Sample size	λ_0	λ_1	Est. τ	Est. λ_0	Est. λ_1
MLE	500	0.2	0.15	3.3936	0.1918	0.1491
		0.25	0.15	4.0992	0.2508	0.1526
		0.3	0.1	4.1629	0.279	0.0999
	1000	0.2	0.15	3.6117	0.198	0.1529
		0.25	0.15	4.3233	0.2356	0.1521
		0.3	0.1	4.4713	0.2744	0.1058
	1500	0.2	0.15	4.3522	0.1924	0.1445
		0.25	0.15	4.6575	0.2385	0.1576
		0.3	0.1	4.9384	0.2819	0.0991
LSE	500	0.2	0.15	5.7787	0.1967	0.1348
		0.25	0.15	5.5386	0.1969	0.1388
		0.3	0.1	4.8569	0.298	0.0835
	1000	0.2	0.15	4.0533	0.2031	0.1523
		0.25	0.15	5.2591	0.2417	0.1414
		0.3	0.1	5.0279	0.3168	0.0932
	1500	0.2	0.15	5.4508	0.1997	0.1378
		0.25	0.15	4.5053	0.2001	0.1522
		0.3	0.1	4.9888	0.2954	0.0991

Application: Clinical Trial of Prometa Treatment Program

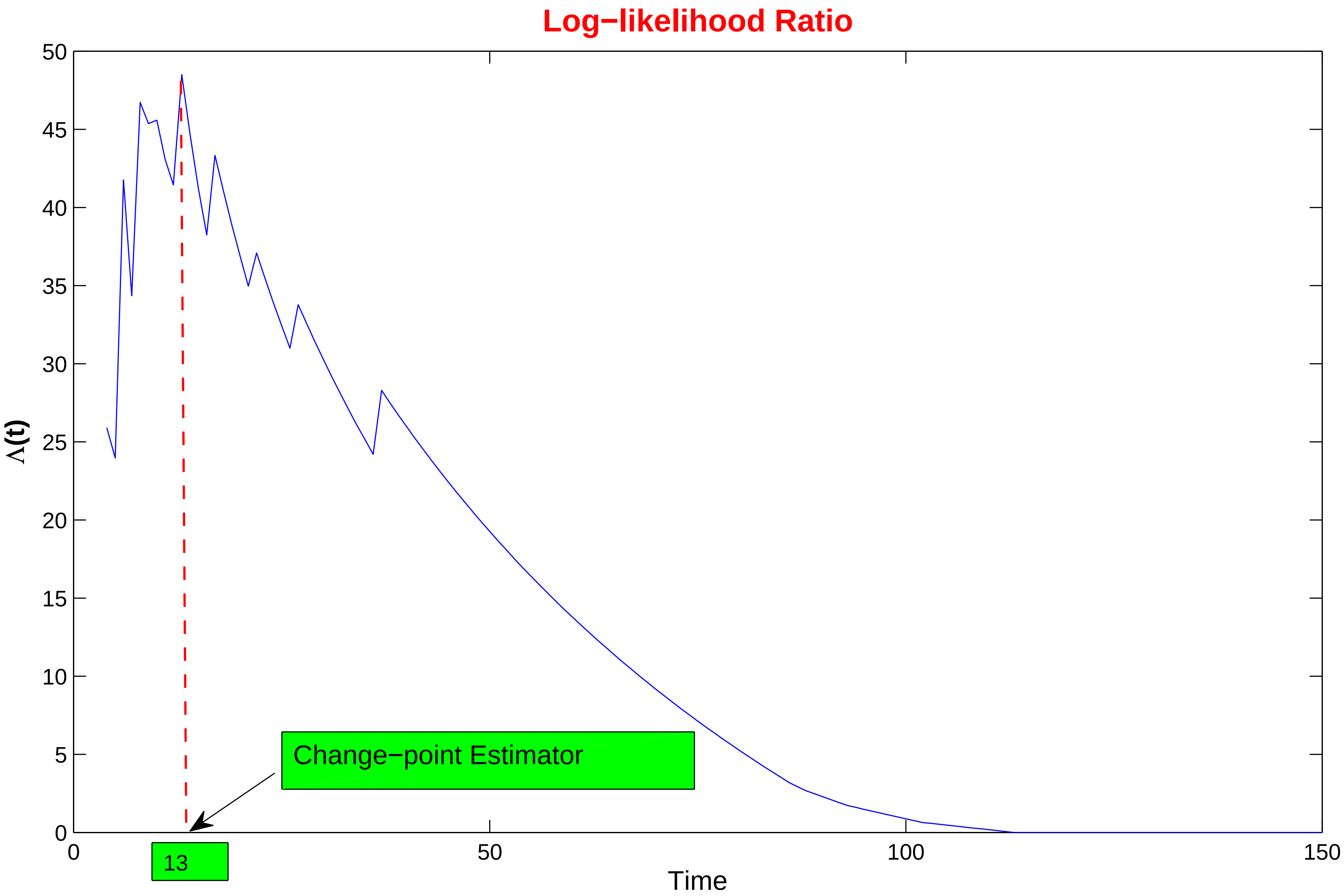
- Patients: heavy methamphetamine-dependent drug users
- Treatment: Prometa Program
- Survival Time: time to relapse(the longer the better)
- Experimental Design:



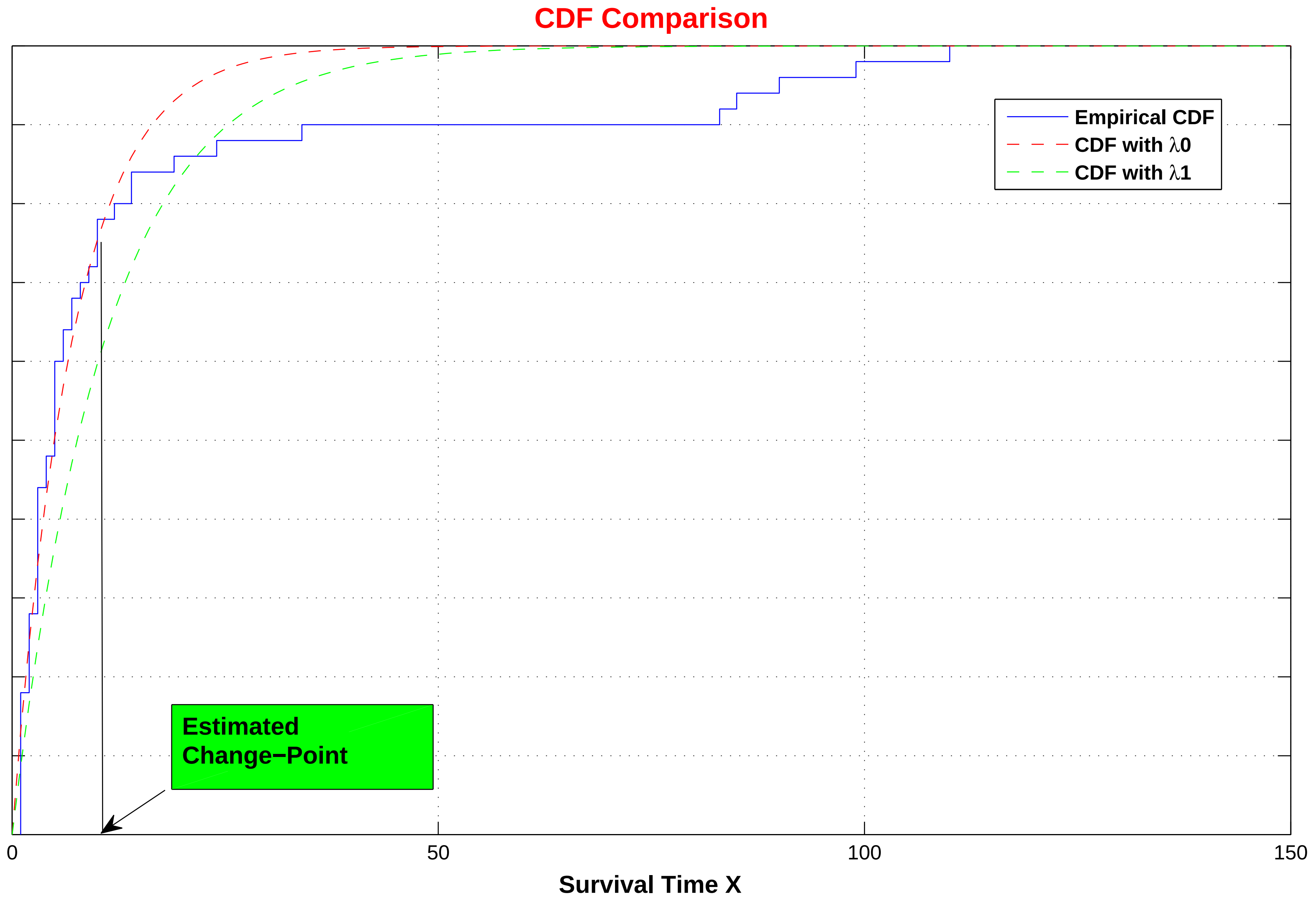
Goal: Estimate change-points after the first 3 treatments.

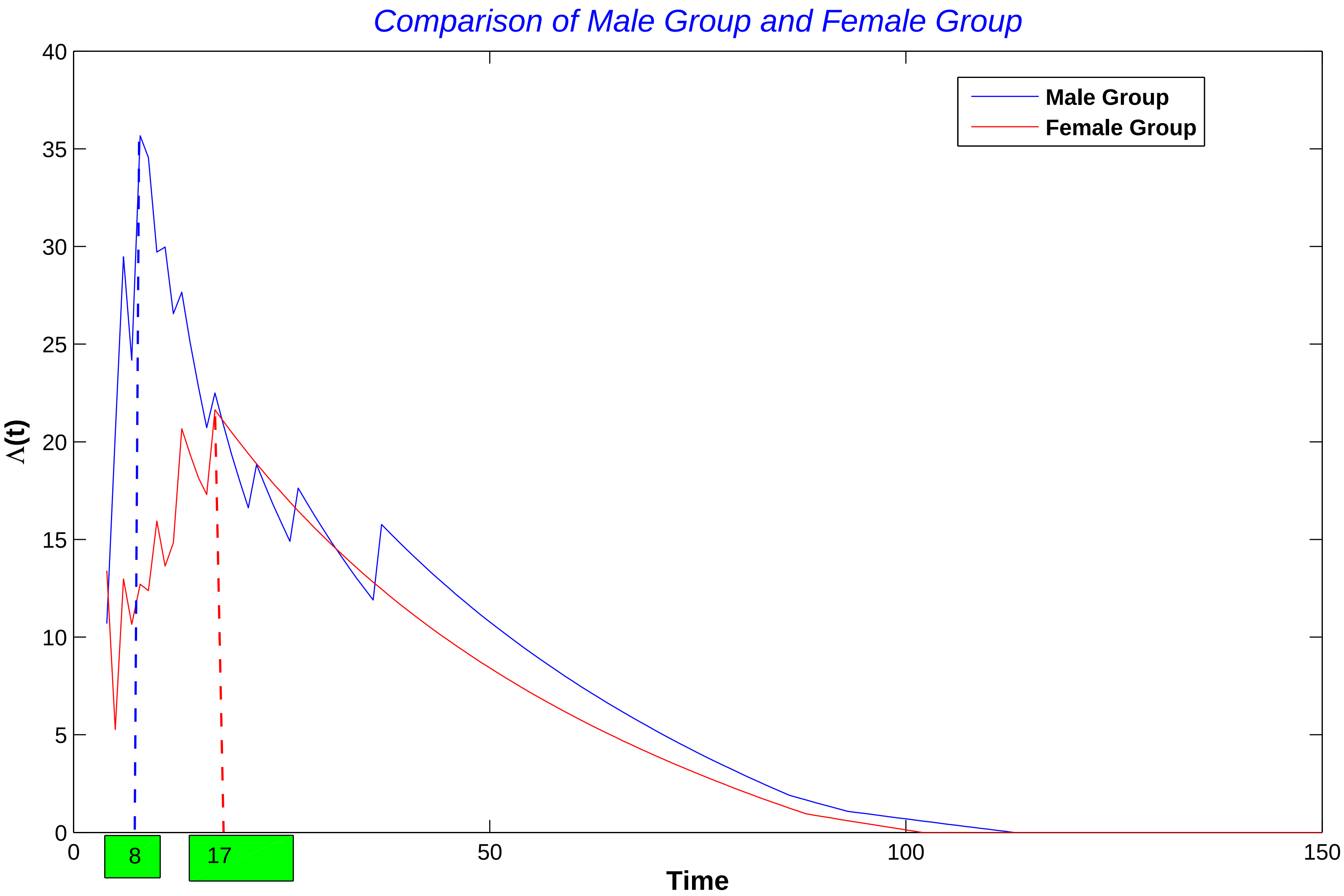


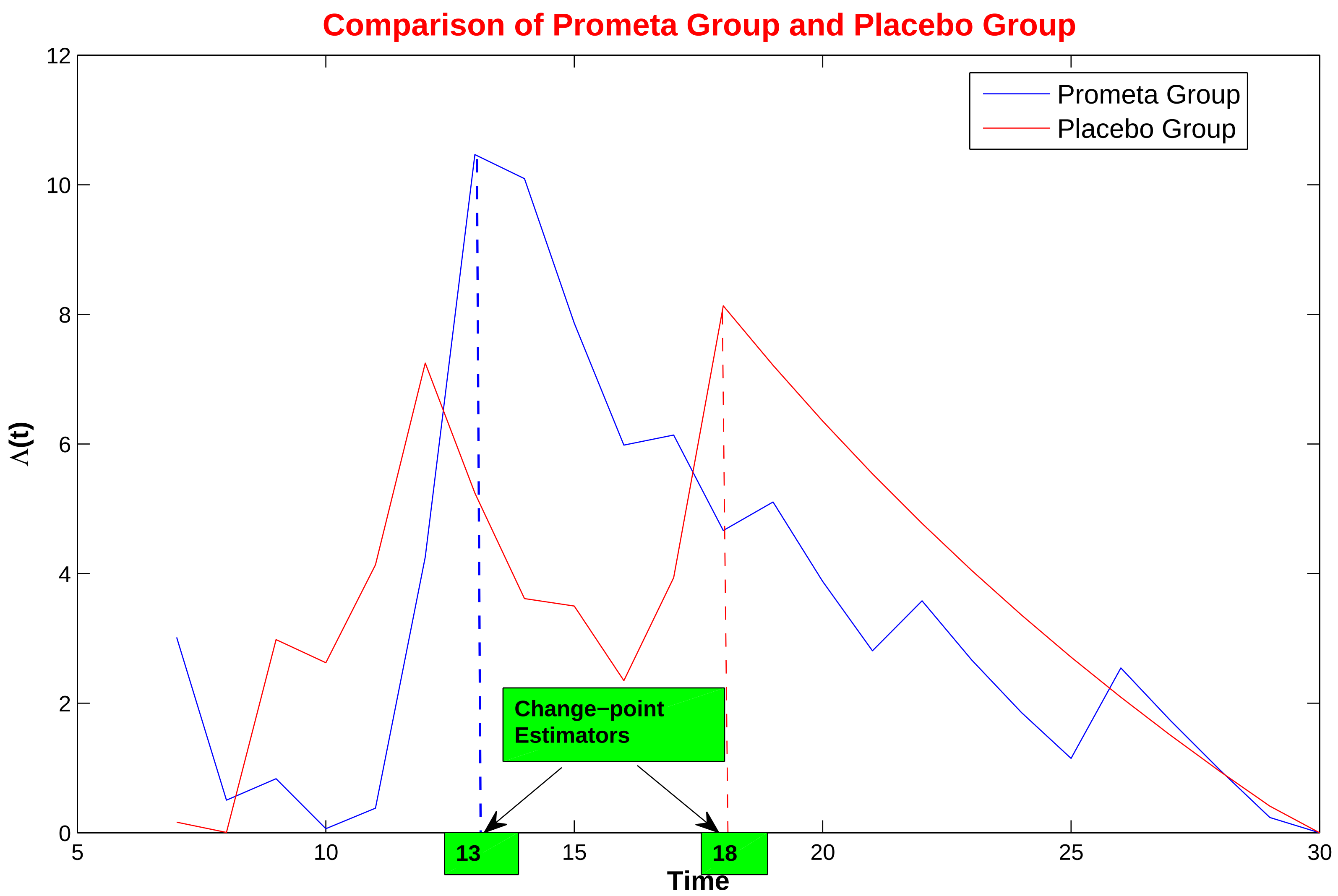
Empirical CDF
F(x)
Survival Time X
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
20
40
60
80
100
120



Log−likelihood Ratio
Λ(t)
50
45
40
35
30
25
20
15
10
5
0
0
50
100
150
Time
Change−point Estimator
13


CDF Comparison
Empirical CDF
CDF with λ0
CDF with λ1
Estimated Change−Point
0
50
100
150
Survival Time X


Comparison of Male Group and Female Group
Male Group
Female Group
Λ(t)
Time
0
5
10
15
20
25
30
35
40
50
100
150
8
17


Comparison of Prometa Group and Placebo Group
Prometa Group
Placebo Group
Λ(t)
Time
0
2
4
6
8
10
12
5
10
15
20
25
30
Change−point
Estimators
13
18

Thank you!