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

\[ X = (X_1, \ldots, X_\tau, X_{\tau+1}, \ldots, X_n) = (X_1, X_2) \]

\[ X_1, \ldots, X_\tau \text{ from } f(x) \]
\[ X_{\tau+1}, \ldots, X_n \text{ from } g(x) \]

where \( \tau = \text{change-point parameter} \)

- Application
  - quality control / epidemiology / financial data

- Theory: MLE, LRT, CUSUM, Bayesian, Exponentially Weighted Moving Average

- Maximum Likelihood Estimate
  - MLE of \( \tau \) is not consistent
  - MLE of \( \tau \) is asymptotically the uniform minimum variance unbiased estimator
Survival Data with a Change-Point

- Piece-wise Constant Hazard Model

\[
\lambda(x) = \lambda_0 \mathbb{1}_{x \leq \tau} + \lambda_1 \mathbb{1}_{x > \tau}
\] (1)

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

\[
f(x_1, \ldots, x_n | \lambda_0, \lambda_1, \tau) = \lambda(x) \exp \left\{ - \int_0^x \lambda(t) dt \right\}
= \prod_{x_i \leq t} (\lambda_0 e^{-\lambda_0 x_i} \mathbb{1}_{x_i \leq \tau} + \lambda_1 e^{-\lambda_0 \tau} - \lambda_1 (x_i - \tau) \mathbb{1}_{x_i > \tau})
\prod_{x_i > t} (e^{-\lambda_0 t} \mathbb{1}_{t \leq \tau} + e^{-\lambda_0 \tau} - \lambda_1 (t - \tau) \mathbb{1}_{t > \tau})
\]
**Maximum Likelihood Estimate**  
(*Without Nuisance Parameters*)

Maximize

\[
\Lambda_t = \log \frac{f(x_1, \ldots, x_n|\lambda_0, \lambda_1, \tau)}{f(x_1, \ldots, x_n|\lambda_0, \tau = \infty)}
\]

\[
= \sum_{i \leq n} y_i
\]

where

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

**Note:** \(\Lambda_t\) is *piecewise linear* in \(\tau\) and maximum of \(\Lambda_t\) should achieve on some \(X_i\).

\[
\hat{\tau} = \left\{ X_{(k)} : \max_k \sum_{i=k+1}^{n} y(i) \right\}
\]
**Convergence Rate of \( \hat{\tau} \)**

**Theorem 1.** Suppose that the true value of change-point is \( \tau_0 \). Then the log-likelihood ratio \( \Lambda_t(\tau) \) has the following properties

(i) \( E\Lambda_t(\tau) \) is strictly increasing in \( \tau \) when \( \tau < \tau_0 \).

(ii) \( E\Lambda_t(\tau) \) is strictly decreasing in \( \tau \) when \( \tau_0 < \tau < t \).

(iii) \( \Lambda_t(\tau) \) is constant after time \( t \).

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

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

where

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

\[
\rho = \lambda_1/\lambda_0.
\]
Confidence Region and its Asymptotic Behavior

\[ R_c = \{ \tau : \max_{0 \leq s \leq t} \Lambda_t(s) - \Lambda_t(\tau) < c \} \] minimizes risk function:
\[ \lambda \mathbb{E}|R_c| + P(\tau \notin R_c) \]
**Theorem 3.** (Asy. behavior of \( P(\tau \in R_c) \)) For \( c \to \infty \)

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

where

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

\[
\pi_1 = \exp \left\{ - \sum_{k=1}^{\infty} k^{-1} P(S'_k > 0) \right\}
\]

\[
\rho_0 = -E_0 W'_m = \left( \frac{\lambda_1}{\lambda_0} - 1 \right) - \log \frac{\lambda_1}{\lambda_0}
\]

\[
\rho_1 = -E_1 W_m = \left( \frac{\lambda_0}{\lambda_1} - 1 \right) - \log \frac{\lambda_0}{\lambda_1}.
\]

**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

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

where $b$ is a constant term.

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

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

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

\[ R_{c_0,c_1} = R_{c_0} \cup R_{c_1} \]
\[ = \{ \tau \leq \hat{\tau} : \Lambda_t(\hat{\tau}) - \Lambda_t(\tau) < c_0 \} \cup \{ \tau > \hat{\tau} : \Lambda_t(\hat{\tau}) - \Lambda_t(\tau) < c_1 \} \]

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

\[
\mathbb{E}W(\tau, R_{c_0,c_1}) \sim \lambda \left( \frac{c_0}{\rho_0} + \frac{c_1}{\rho_1} + b + o(1) \right) + \pi_0 \pi_1 \left( \frac{e^{-c_0}}{\rho_0} + \frac{e^{-c_1}}{\rho_1} \right)
\]
\[ = (\lambda c_0 + e^{-c_0} \pi_0 \pi_1) \frac{1}{\rho_0} + (\lambda c_1 + e^{-c_1} \pi_0 \pi_1) \frac{1}{\rho_1} + \lambda(b + o(1)) \]

**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

\[ \hat{\lambda}_0 = R_T \left( \sum (X_i 1_{X_i \leq \tau < t}) + \tau(n - R_T) \right)^{-1} \]

\[ \hat{\lambda}_1 = R_t - R_T \left( \sum X_i \leq t (X_i - \tau)^+ + (t - \tau)(n - R_t) \right)^{-1} \]

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

- 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 \( S(t) \):

\[
\tilde{S}_n(x) = \prod_{x(j) \leq x} \left( \frac{n - j}{n - j + 1} \right)^{\delta(j)}
\]  

(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

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

(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

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

for sufficiently large $n$. 
Comparison of MLE and LSE through Simulation Study

- Simulate data from model (1)
- $\tau = 5, t = 20$
- $(\lambda_0, \lambda_1) = (0.2, 0.15), (0.25, 0.15), \text{ and } (0.3, 0.1)$.
- Samples sizes from 500, 1000, 1500 were considered.
- Conclusions
  - Both MLE and LSE are more accurate when we increase the difference between $\lambda_0$ and $\lambda_1$ for the same sample size.
  - Both MLE and LSE converge to the true change-point when we increase the sample size for the same $\lambda_0$ and $\lambda_1$.
  - LSE is more accurate than MLE for the same sample size and same failure rates.
<table>
<thead>
<tr>
<th>Method</th>
<th>Sample size</th>
<th>$\lambda_0$</th>
<th>$\lambda_1$</th>
<th>Est. $\tau$</th>
<th>Est. $\lambda_0$</th>
<th>Est. $\lambda_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>500</td>
<td>0.2</td>
<td>0.15</td>
<td>3.3936</td>
<td>0.1918</td>
<td>0.1491</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.25</td>
<td>0.15</td>
<td>4.0992</td>
<td>0.2508</td>
<td>0.1526</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3</td>
<td>0.1</td>
<td>4.1629</td>
<td>0.279</td>
<td>0.0999</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.2</td>
<td>0.15</td>
<td>3.6117</td>
<td>0.198</td>
<td>0.1529</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.25</td>
<td>0.15</td>
<td>4.3233</td>
<td>0.2356</td>
<td>0.1521</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3</td>
<td>0.1</td>
<td>4.4713</td>
<td>0.2744</td>
<td>0.1058</td>
</tr>
<tr>
<td></td>
<td>1500</td>
<td>0.2</td>
<td>0.15</td>
<td>4.3522</td>
<td>0.1924</td>
<td>0.1445</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.25</td>
<td>0.15</td>
<td>4.6575</td>
<td>0.2385</td>
<td>0.1576</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3</td>
<td>0.1</td>
<td>4.9384</td>
<td>0.2819</td>
<td>0.0991</td>
</tr>
<tr>
<td>LSE</td>
<td>500</td>
<td>0.2</td>
<td>0.15</td>
<td>5.7787</td>
<td>0.1967</td>
<td>0.1348</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.25</td>
<td>0.15</td>
<td>5.5386</td>
<td>0.1969</td>
<td>0.1388</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3</td>
<td>0.1</td>
<td>4.8569</td>
<td>0.298</td>
<td>0.0835</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.2</td>
<td>0.15</td>
<td>4.0533</td>
<td>0.2031</td>
<td>0.1523</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.25</td>
<td>0.15</td>
<td>5.2591</td>
<td>0.2417</td>
<td>0.1414</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3</td>
<td>0.1</td>
<td>5.0279</td>
<td>0.3168</td>
<td>0.0932</td>
</tr>
<tr>
<td></td>
<td>1500</td>
<td>0.2</td>
<td>0.15</td>
<td>5.4508</td>
<td>0.1997</td>
<td>0.1378</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.25</td>
<td>0.15</td>
<td>4.5053</td>
<td>0.2001</td>
<td>0.1522</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3</td>
<td>0.1</td>
<td>4.9888</td>
<td>0.2954</td>
<td>0.0991</td>
</tr>
</tbody>
</table>
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:

  | Visit1 | Trt1 | Trt2 | Trt3 | Visit5 | Visit6 | Trt4 | Trt5 | Visit7 | ... | Visit17 |
  | Day1   | Day2 | Day3 | Day4 | 10     | 17     | 23   | 24   | 30     | 84   |

*Goal:* Estimate change-points after the first 3 treatments.
Comparison of Male Group and Female Group

Male Group
Female Group

\( \Lambda(t) \)
Comparison of Prometa Group and Placebo Group

Prometa Group
Placebo Group

Change-point Estimators

Time

\( \Lambda(t) \)
Thank you!