Survival Analysis for Interval Censored Data

- Nonparametric Estimation

Seminar of Statistics, ETHZ
Group 8, 02. May 2011
Martina Albers, Nanina Anderegg, Urs Müller
Overview

• Examples
• Nonparametric MLE for Current Status Censored Data
  – Asymptotic Behavior?
INTRODUCTION

Goals
Different types of censoring
Terminology
Examples
Censored data

- Right censoring
- Current status censoring
- Interval censoring case $k$
- Mixed case interval censoring
- Bivariate interval censored data
Right censoring

1. $X$: failure time
   - $X \sim F$
2. $C$: censoring time
   - $C \sim G$
3. $C$ independent of $X$
4. $n$ observations: iid copies of

\[
(U, \Delta) \equiv \left( \min(X, C), 1_{(X \leq C)} \right)
\]

Goal: Estimate distribution function $F(x) = P[X \leq x]$
Censored data

• Right censoring
• Current status censoring
• Interval censoring case $k$
• Mixed case interval censoring
• Bivariate interval censored data
Current status censoring

1. \( X \): failure time
   - \( X \sim F \)
2. \( T \): observation time
   - \( T \sim G \)
3. \( X \) independent of \( T \)
4. \( n \) observations: iid copies of
   \[
   (T, \Delta) \equiv (T, 1_{\{X \leq T\}})
   \]

Goal: Estimate distribution function \( F(x) = P[X \leq x] \)
Right censoring

vs.

Current status censoring

Right Censoring
• Observe:

\((U, \Delta) \equiv \left( \min(X, C), 1_{\{X \leq C\}} \right)\)

Current Status Censoring
• Observe:

\((T, \Delta) \equiv \left( T, 1_{\{X \leq T\}} \right)\)
Example: Lung cancer data

• Time of onset of lung cancer (non lethal)
• 144 (special) mice (likely to get cancer)
  – sacrifice each mouse at random time
    → Determine current status: cancer/no cancer
  – 2 groups
    • Conventional environment (96)
    • Germfree environment (48)

• Time to onset of lung cancer:
  [0 (start) → onset of cancer]
Time to onset of cancer

• At time of sacrifice: find tumor
  → Onset occurred at some earlier point
• At time of sacrifice: no tumor
  → Onset will occur later/never

→ Goal: estimate distribution of time to onset!
  (analyze for the two groups separately)
Mathematical Formulation

• $X$: time to onset of lung cancer

• $T$: time of sacrifice
  – Random time! (chosen randomly!)

• Both are measured from the beginning of study
Mathematical Formulation

• \((x, t, \delta)\) realization of \((X, T, \Delta)\)
• Cancer was observed at time of sacrifice:
  – \(x \leq t\)
  – observe: \((t, 1_{\{x \leq t\}}) = (t, 1)\)
  – conclusion: \(x \in (0, t]\)
• Cancer not observed at time of sacrifice:
  – \(x > t\)
  – observe: \((t, 1_{\{x \leq t\}}) = (t, 0)\)
  – conclusion: \(x \in (t, \infty)\)
Right censoring

vs.

Current status censoring

Right Censoring

• Observe:

\[(U, \Delta) \equiv \left( \min(X, C), 1_{(X \leq C)} \right)\]

• Either know \( x \) exactly
  or \( x \in (c, \infty) \)

Current Status Censoring

• Observe:

\[(T, \Delta) \equiv \left( T, 1_{\{X \leq T\}} \right)\]

• NEVER know \( x \) exactly, i.e.
  \( x \in (0, t] \) or \( x \in (t, \infty) \)
Right censoring

$x \in (c, \infty)$
Current status censoring

(*Interval censoring case 1*)

\[ x \in (0, t] \quad \text{and} \quad x \in (t, \infty) \]
### The data

#### Table 1.3. Death times in days for 144 male RFM mice with lung tumors

<table>
<thead>
<tr>
<th>Group</th>
<th>Tumor status</th>
<th>Death times</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE</td>
<td>With tumor</td>
<td>381, 477, 485, 515, 539, 563, 565, 582, 603, 616, 624, 650, 651, 656, 659, 672, 679, 698, 702, 709, 723, 731, 775, 779, 795, 811, 839</td>
</tr>
<tr>
<td>GE</td>
<td>With tumor</td>
<td>546, 609, 692, 692, 710, 752, 773, 781, 782, 789, 808, 810, 814, 842, 846, 851, 871, 873, 876, 888, 888, 890, 894, 896, 911, 913, 914, 914, 916, 921, 921, 926, 936, 945, 1008</td>
</tr>
<tr>
<td></td>
<td>No tumor</td>
<td>412, 524, 647, 648, 695, 785, 814, 817, 851, 880, 913, 942, 986</td>
</tr>
</tbody>
</table>

**Conventional environment (96)**

**Germfree environment (48)**
The data

• With tumor: 381
  ➔ After 381 days: tumor was found in mouse

• No tumor: 45
  ➔ After 45 days: no tumor was found in mouse
Censored data

• Right censoring
• Current status censoring
• Interval censoring case $k$
• Mixed case interval censoring
• Bivariate interval censored data
2 intervals $\Rightarrow (k+1)$ intervals!
Theoretical example: HIV infection

• Age at HIV infection

• Representative sample
  – Follow each person for 2 years
  – 2 HIV tests during this period
    → Determine: negative/positive

• Time to infection:
  [0 (start) → HIV infection]
Interval censoring case $k=2$

1. $X$: failure time
   - $X \sim F$
2. $(T_1, T_2)$: observation times
   - $(T_1, T_2) \sim G$
3. $X$ independent of $(T_1, T_2)$
4. $n$ observations: iid copies of
   \[
   (\bar{T}, \bar{\Delta}) \equiv \left((T_1, T_2), (1_{X\leq T_1}, 1_{T_1<X\leq T_2}, 1_{T_2<X})\right)
   \]

Goal: Estimate distribution function $F(x) = P[X \leq x]$
2 observation times

\[ \Delta_1 \quad T_1 \quad \Delta_2 \quad T_2 \quad (\Delta_3) \]

Nota bene: \[ \Delta_3 = 1 - \Delta_1 - \Delta_2 \]
Interval Censoring, case 1 vs. Interval Censoring, case $k$

**Case 1**
- 1 observation per subject:

$$(T, \Delta) \equiv (T, 1_{\{X \leq T\}})$$

**Case $k$**
- $k$ observations per subject:

$$(\tilde{T}, \tilde{\Delta}) = 
\left( (T_1, T_2, \ldots, T_k), (1_{\{X \leq T_1\}}, 1_{\{T_1 < X \leq T_2\}}, \ldots, 1_{\{T_k < X\}}) \right)$$

$$(T, \Delta) = (T, 1_{\{X \leq T\}}, 1_{\{T < X\}})$$
Mathematical Formulation

- $X$: age at HIV infection

- $(T_1, T_2)$: ages at time of HIV tests
  - Random times (chosen randomly!)

- All times are measured from the beginning of study
Mathematical Formulation

- \((x, t_1, t_2, \delta_1, \delta_2, \delta_3)\) realization of \((X, T_1, T_2, \Delta_1, \Delta_2, \Delta_3)\)
- 1\(^{st}\) test is positive:
  - \(x \leq t_1\)
  - observe: \((t_1, t_2, 1_{\{x \leq t_1\}}, 1_{\{t_1 < x \leq t_2\}}, 1_{\{t_2 < x\}}) = (t_1, t_2, 1, 0, 0)\)
  - conclusion: \(x \in (0, t_1]\)
- 1\(^{st}\) test is negative, 2\(^{nd}\) test is positive:
  - \(t_1 < x \leq t_2\)
  - observe: \((t_1, t_2, 1_{\{x \leq t_1\}}, 1_{\{t_1 < x \leq t_2\}}, 1_{\{t_2 < x\}}) = (t_1, t_2, 0, 1, 0)\)
  - conclusion: \(x \in [t_1, t_2)\)
- 1\(^{st}\) and 2\(^{nd}\) test are negative:
  - \(x > t_2\)
  - observe: \((t_1, t_2, 1_{\{x \leq t_1\}}, 1_{\{t_1 < x \leq t_2\}}, 1_{\{t_2 < x\}}) = (t_1, t_2, 0, 0, 1)\)
  - conclusion: \(x \in [t_2, \infty)\)
Censored data

• Right censoring
• Current status censoring
• Interval censoring case \( k \)
• **Mixed case interval censoring**
• Bivariate interval censored data
fixed $k \rightarrow$ variable $K$!

• for all patients: $k$ observations (fix!)

• Let number of observations vary from patient to patient
  – Define random variable $K$ as number of visits
HIV infection

• Patient A: $K = 2$
  $\left(t_1, t_2, 1_{\{x \leq t_1\}}, 1_{\{t_1 < x \leq t_2\}}, 1_{\{t_2 < x\}}\right) = (t_1, t_2, 1, 0, 0)$

• Patient B: $K = 3$
  $\left(t_1, t_2, t_3, 1_{\{x \leq t_1\}}, 1_{\{t_1 < x \leq t_2\}}, 1_{\{t_2 < x \leq t_3\}}, 1_{\{t_3 < x\}}\right) = (t_1, t_2, t_3, 0, 0, 1, 0)$

• Patient C: $K = 1$
  $\left(t_1, 1_{\{x \leq t_1\}}, 1_{\{t_1 < x\}}\right) = (t_1, 0, 1)$
Example: Breast cosmesis data

- Time to breast retraction
- 94 patients
  - clinic visits to determine current status retraction/no retraction
    - Number of visits vary from patient to patient!
  - 2 groups
    - Radiotherapy (46)
    - Radiotherapy + chemotherapy (48)
- Time to breast retraction:
  [0 (start) → retraction]
### Table 4

*Interval of cosmetic deterioration (retraction) for early breast cancer patients treated with radiotherapy and chemotherapy vs. radiotherapy alone*

<table>
<thead>
<tr>
<th></th>
<th>Radiotherapy</th>
<th>Radiotherapy and Chemotherapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>(45, ___)*</td>
<td>(25, 37)</td>
<td>(8, 12)</td>
</tr>
<tr>
<td>(6, 10)</td>
<td>(46, ___)</td>
<td>(0, 5)</td>
</tr>
<tr>
<td>(0, 7)</td>
<td>(26, 40)</td>
<td>(24, 31)</td>
</tr>
<tr>
<td>(46, ___)</td>
<td>(46, ___)</td>
<td>(17, 27)</td>
</tr>
<tr>
<td>(46, ___)</td>
<td>(27, 34)</td>
<td>(17, 23)</td>
</tr>
<tr>
<td>(7, 16)</td>
<td>(36, 44)</td>
<td>(24, 30)</td>
</tr>
<tr>
<td>(17, ___)</td>
<td>(46, ___)</td>
<td>(17, 23)</td>
</tr>
<tr>
<td>(7, 14)</td>
<td>(36, 48)</td>
<td>(17, 25)</td>
</tr>
<tr>
<td>(37, 44)</td>
<td>(37, ___)</td>
<td>(13, ___)</td>
</tr>
<tr>
<td>(0, 8)</td>
<td>(40, ___)</td>
<td>(11, 13)</td>
</tr>
<tr>
<td>(4, 11)</td>
<td>(17, 25)</td>
<td>(13, ___)</td>
</tr>
<tr>
<td>(15, ___)</td>
<td>(46, ___)</td>
<td>(17, 26)</td>
</tr>
<tr>
<td>(11, 15)</td>
<td>(11, 18)</td>
<td>(17, 26)</td>
</tr>
<tr>
<td>(22, ___)</td>
<td>(38, ___)</td>
<td>(32, ___)</td>
</tr>
<tr>
<td>(46, ___)</td>
<td>(5, 12)</td>
<td>(44, 48)</td>
</tr>
<tr>
<td>(46, ___)</td>
<td>(36, ___)</td>
<td>(14, 17)</td>
</tr>
</tbody>
</table>

* A right endpoint “___” indicates observation is right-censored.
The data

- \((6,10]\)
  At 6 months: patient showed no deterioration in initial cosmetic result
  By 10 months: retraction was present

- \((45, \_]\)
  At 45 months: patient showed no deterioration in initial cosmetic result
  "\(\_\)" : No second test
Censored data

- Right censoring
- Current status censoring
- Interval censoring case $k$
- Mixed case interval censoring
- **Bivariate interval censored data**
Bivariate interval censoring

1. \((X,Y)\) : failure times
   • \((X,Y) \sim F\)
2. \(\vec{U} = (U_1, U_2), \vec{V} = (V_1, V_2)\) : observation times
   • \((\vec{U}, \vec{V}) \sim G\)
3. \((X,Y)\) independent of \((\vec{U}, \vec{V})\)
4. \(n\) observations: iid copies of \((\vec{U}, \vec{V}, \vec{\Delta})\)

where \(\vec{\Delta} = (\Delta_{11}, \Delta_{12}, \Delta_{13}, \Delta_{21}, \Delta_{22}, \Delta_{23}, \Delta_{31}, \Delta_{32}, \Delta_{33})\)

Goal: Estimate distribution function

\[
F(x, y) = P[X \leq x, Y \leq y]
\]
Bivariate interval censoring

\[ \Delta_{ij} = 1_{\{(X,Y) \in R_{ij}\}} \]

\[ R_{11} = (0,U_1] \times (0,V_1] \]
\[ R_{12} = (U_1,U_2] \times (0,V_1] \]
\[ R_{13} = (U_2,\infty) \times (0,V_1] \]
\[ R_{21} = (0,U_1] \times (V_1,V_2] \]
\[ R_{22} = (U_1,U_2] \times (V_1,V_2] \]
\[ R_{23} = (U_2,\infty) \times (V_1,V_2] \]
\[ R_{31} = (0,U_1] \times (V_2,\infty) \]
\[ R_{32} = (U_1,U_2] \times (V_2,\infty) \]
\[ R_{33} = (U_2,\infty) \times (V_2,\infty) \]
Sketch on blackboard...
Example: ACTG 181

- Time to shedding of a cytomegalovirus (CMV)
- Time to colonization of mycobacterium avium complex (MAC)
- 204 patients
  - test for CMV shedding and MAC colonization at least once during trial (no prior CMV/MAC diagnosis)
    → Tested at regular monthly intervals
Time to CMV shedding and MAC colonization

- Patient didn’t miss any clinical visit:
  - record month of $1^{st}$ positive test
    - (discrete failure time data)
- Patient missed some visits and was tested positive after missed visit:
  - record time interval
    - (discrete interval censored failure time data)

Goal: estimate distribution of time of shedding and of colonization!
Mathematical Formulation

• $X$: time to CMV shedding
• $Y$: time to MAC colonization
• $(U_1, U_2)$: times of tests for CMV shedding
• $(V_1, V_2)$: times of tests for MAC colonization

• Model is not perfect here!
  – Number of observations varies from patient to patient
  – Model should include different or random number of observation times
The data

- 204 subjects:
  - 68 interval censored time of CMV shedding
  - 10 interval censored time of MAC colonization
  - 89 right censored times of CMV shedding
  - 190 right censored times of MAC colonization
  - 46 left censored times of CMV shedding
  - 4 left censored times of MAC colonization

- Test every 12 weeks, i.e. quarterly screens
- Visit rounded to closest quarter
The data – an excerpt...

<table>
<thead>
<tr>
<th>I × R</th>
<th>#</th>
<th>R × R</th>
<th>#</th>
<th>L × R</th>
<th>#</th>
<th>R × I</th>
<th>#</th>
<th>R × L</th>
<th>#</th>
<th>I × I</th>
<th>#</th>
<th>L × L</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0, 3) × 0</td>
<td>(3)</td>
<td>0 × 0</td>
<td>(6)</td>
<td>0 × 0</td>
<td>(9)</td>
<td>0 × (0, 3)</td>
<td>(1)</td>
<td>3 × 0</td>
<td>(1)</td>
<td>(0, 3) × (0, 6)</td>
<td>(1)</td>
<td>0 × 0</td>
<td>(1)</td>
</tr>
<tr>
<td>(0, 3) × 3</td>
<td>(1)</td>
<td>3 × 0</td>
<td>(2)</td>
<td>0 × 3</td>
<td>(3)</td>
<td>6 × (0, 6)</td>
<td>(1)</td>
<td>9 × 0</td>
<td>(1)</td>
<td>(3, 6) × (6, 12)</td>
<td>(1)</td>
<td>0 × 0</td>
<td>(1)</td>
</tr>
<tr>
<td>(0, 3) × 6</td>
<td>(3)</td>
<td>6 × 0</td>
<td>(1)</td>
<td>0 × 6</td>
<td>(10)</td>
<td>6 × (6, 6)</td>
<td>(1)</td>
<td>12 × 0</td>
<td>(1)</td>
<td>(9, 9) × (9, 9)</td>
<td>(1)</td>
<td>0 × 0</td>
<td>(1)</td>
</tr>
<tr>
<td>(0, 6) × 6</td>
<td>(1)</td>
<td>6 × 3</td>
<td>(2)</td>
<td>0 × 9</td>
<td>(6)</td>
<td>12 × (0, 3)</td>
<td>(1)</td>
<td>12 × (0, 6)</td>
<td>(1)</td>
<td>0 × 0</td>
<td>(1)</td>
<td>0 × 0</td>
<td>(1)</td>
</tr>
<tr>
<td>(0, 3) × 9</td>
<td>(1)</td>
<td>6 × 6</td>
<td>(3)</td>
<td>0 × 12</td>
<td>(8)</td>
<td>12 × (0, 6)</td>
<td>(1)</td>
<td>15 × (0, 3)</td>
<td>(1)</td>
<td>0 × 0</td>
<td>(1)</td>
<td>0 × 0</td>
<td>(1)</td>
</tr>
<tr>
<td>(0, 3) × 12</td>
<td>(5)</td>
<td>6 × 9</td>
<td>(1)</td>
<td>0 × 15</td>
<td>(5)</td>
<td>21 × (15, 15)</td>
<td>(1)</td>
<td>0 × 0</td>
<td>(1)</td>
<td>0 × 0</td>
<td>(1)</td>
<td>0 × 0</td>
<td>(1)</td>
</tr>
<tr>
<td>(0, 3) × 15</td>
<td>(5)</td>
<td>9 × 0</td>
<td>(2)</td>
<td>0 × 18</td>
<td>(4)</td>
<td>0 × 0</td>
<td>(1)</td>
<td>0 × 0</td>
<td>(1)</td>
<td>0 × 0</td>
<td>(1)</td>
<td>0 × 0</td>
<td>(1)</td>
</tr>
<tr>
<td>(0, 6) × 15</td>
<td>(1)</td>
<td>9 × 9</td>
<td>(3)</td>
<td>0 × 21</td>
<td>(1)</td>
<td>0 × 0</td>
<td>(1)</td>
<td>0 × 0</td>
<td>(1)</td>
<td>0 × 0</td>
<td>(1)</td>
<td>0 × 0</td>
<td>(1)</td>
</tr>
<tr>
<td>(3, 3) × 3</td>
<td>(1)</td>
<td>9 × 12</td>
<td>(1)</td>
<td>0 × 0</td>
<td>(1)</td>
<td>0 × 0</td>
<td>(1)</td>
<td>0 × 0</td>
<td>(1)</td>
<td>0 × 0</td>
<td>(1)</td>
<td>0 × 0</td>
<td>(1)</td>
</tr>
</tbody>
</table>

Number of patients
The data

• \((0,6) \times 15\):
  - at start: no CMV shedding
  - missed controls in the following 6 months
  - after 6 months: first CMV shedding
  - MAC colonization after 15 months

• \((3,3)\):
  - actual left and right endpoints are rounded to the same number.
Derive the likelihood for different censoring models

THE LIKELIHOOD
Nonparametric MLE

• Completely nonparametric
  ➔ Do not assume that failure time distribution follows any parametric model

• Only assumption: variable of interest \( X \) is independent of censoring resp. observation times \( C \) resp. \( T \)
Recall: Right censored data

Likelihood of $n$ iid observations $(u_1, \delta_1), (u_2, \delta_2), \ldots, (u_n, \delta_n)$

$$
\prod_{i=1}^{n} f(u_i)^{\delta_i} (1 - F(u_i))^{1-\delta_i} (1 - G(u_i -))^\delta_i g(u_i)^{1-\delta_i}
$$

→ Nonparametric MLE $F$ maximizes:

$$
L_n(F) = \prod_{i=1}^{n} F(\{u_i\})^{\delta_i} (1 - F(u_i))^{1-\delta_i}
$$

where $F(\{u\}) = F(u) - F(u-)$
Other representation: observed sets

- \( R_1, R_2, \ldots, R_n \): observed sets
- Contain the unobservable realization of \( X \)
- i.e.:
  \[
  R_i = \begin{cases} 
  \{ u_i \} & \text{if} \quad \delta_i = 1 \\
  (u_i, \infty) & \text{if} \quad \delta_i = 0
  \end{cases}
  \]

\[ \Rightarrow \text{Nonparametric MLE } F \text{ maximizes:} \]

\[
L_n (F) = \prod_{i=1}^{n} F (\{u_i\})^{\delta_i} (1 - F (u_i))^{1-\delta_i} = \prod_{i=1}^{n} P_F (R_i)
\]

Where \( P_F (R_i) \) is the probability under distribution \( F \) that \( X \in R_i \)
Current status data

• Recall: n observations: iid copies of

\[ (T, \Delta) \equiv (T, 1_{\{X \leq T\}}) \]

i.e.: \((t_1, \delta_1), (t_2, \delta_2), \ldots, (t_n, \delta_n)\)

• \(X \sim F\)

• \(T \sim G\)

• Likelihood = product of densities of observations

• Consider \(\Delta = 0\) and \(\Delta = 1\) separately, then combine

\textit{cf. Blackboard!}
Current status data

Likelihood of \( n \) iid observations \((t_1, \delta_1), (t_2, \delta_2), \ldots, (t_n, \delta_n)\)

\[
\prod_{i=1}^{n} F(t_i)^{\delta_i} (1 - F(t_i))^{1 - \delta_i} g(t_i)
\]

\(\rightarrow\) Nonparametric MLE \( F \) maximizes:

\[
L_n(F) = \prod_{i=1}^{n} F(t_i)^{\delta_i} (1 - F(t_i))^{1 - \delta_i}
\]
Other representation: observed sets

- $R_1, R_2, ..., R_n$ : observed sets
- Recall: contain the unobservable realization of $X$
- i.e.:
  \[
  R_i = \begin{cases} 
  (0, t_i] & \text{if } \delta_i = 1 \\
  (t_i, \infty) & \text{if } \delta_i = 0 
  \end{cases}
  \]

$\Rightarrow$ Nonparametric MLE $F$ maximizes:

\[
L_n (F) = \prod_{i=1}^{n} F(t_i)^{\delta_i} (1 - F(t_i))^{1-\delta_i} = \prod_{i=1}^{n} P_F(R_i)
\]

Where $P_F(R_i)$ is the probability under distribution $F$ that $X \in R_i$
General form of the likelihood

• Representation in terms of observed sets:

\[ R_1, R_2, \ldots, R_n \]

\[ \Rightarrow \quad L_n(F) = \prod_{i=1}^{n} P_F(R_i) \]

where \( R_i \) change depending on censoring model!
Reducing the Optimization Problem

• Likelihood

\[ L_n(F) = \prod_{i=1}^{n} P_F(R_i) \]

• Optimization problem:

\[ \sup_{F \in \mathcal{F}} l_n(F), \]

with \( l_n(F) = \log L_n(F) = \sum_{i=1}^{n} \log P_F(R_i) \)

where \( \mathcal{F} \) is the space of all distribution functions on the appropriate space.
Reducing the Optimization Problem

• **Problem**
  – infinite-dimensional optimization problem

• **Solution**
  – maximal intersections $A_1, \ldots, A_m$ of observed sets $R_1, \ldots, R_n$

• **Definition**

  $A_j \neq \emptyset$ is a *maximal intersection* iff

  – $A_j = \bigcap_{i \in \beta_j} R_i$ for some subset $\beta_j \subseteq \{1, \ldots, n\}$

  – no strict superset $\beta_j^* \subseteq \{1, \ldots, n\}$ of $\beta_j$ s.t.: $\bigcap_{i \in \beta_j^*} R_i \neq \emptyset$. 
Reducing the Optimization Problem

• Maximize $\sum_{i=1}^{n} \log P_F(R_i)$,
→ distribute probability mass (total amount $I$) on probability space.

• Properties of MLE
  – No mass outside the sets $R_1, \ldots, R_n$
  – Some positive mass in each set $R_1, \ldots, R_n$
  – Mass only in maximal intersections: $A_1, \ldots, A_m$
  – indifferent to the distribution of mass within the maximal intersections (representational non-uniqueness)
Reducing the Optimization Problem

- \( \alpha_1, \ldots, \alpha_m \) : masses in corresponding maximal intersection
- \( C \): \((m \times n)\) clique matrix, with \( C_{ji} = 1 \{ A_j \subseteq R_i \} \)
- **Likelihood in terms of** \( \alpha \)

\[
l_n(\alpha) = \sum_{i=1}^{n} \log P_{\alpha}(R_i)
= \sum_{i=1}^{n} \log \left( \sum_{j=1}^{n} \alpha_j 1 \{ A_j \subseteq R_i \} \right)
= \sum_{i=1}^{n} \log \left( C^T \alpha \right)_i
\]
Existence and (non-)uniqueness of the MLE

• **Notation** \( l_n(\hat{\alpha}) = \max_{\alpha \in \mathcal{A}} l_n(\alpha) \)

  where

  \[ \mathcal{A} = \left\{ \alpha \in \mathbb{R}^m : \alpha_j \geq 0; j = 1, \ldots, m; \sum_{i=1}^m \alpha_j = 1 \right\} \]

• **Theorem 3.1**
  The MLE \( \hat{\alpha} \) exists.

• **Theorem 3.2**
  Log-likelihood \( \sum_{i=1}^n \log P_F(R_i) \) strictly concave in \( P_F(R) \).

  Thus, MLE estimates the probabilities \( P_F(R_1), \ldots, P_F(R_n) \)

  of the observation rectangles uniquely.

*Proof*: cf. Blackboard
Existence and (non-)uniqueness of the MLE

• Remarks
  – Log-likelihood (not strictly) concave in $F$
  $\Rightarrow$ two different functions $F_1, F_2 \in \mathcal{F}$ can yield same vector $(P_F(R_1), \ldots, P_F(R_n))$
  – The estimation of $F$ is not necessarily unique.
  – Same holds for $\alpha$ (mixture non-uniqueness)
(Non-)uniqueness of the MLE

• Example

\[
\max_{\alpha \in \mathbb{R}^4} \sum_{i=1}^{4} \log P_\alpha(R_i)
\]

\[= \max \left\{ \log (\alpha_1 + \alpha_4) + \log (\alpha_1 + \alpha_2) + \log (\alpha_2 + \alpha_3) + \log (\alpha_3 + \alpha_4) \right\} \]

subject to

- \(\alpha_j \geq 0, \ j = 1, \ldots, 4\)

- \(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1\)

- Solution:

\[
\hat{\alpha}_1 = \hat{\alpha}_3 = \frac{1}{2} - x, \quad \hat{\alpha}_2 = \hat{\alpha}_4 = x, \quad \text{for some } x \in \left[0, \frac{1}{2}\right]
\]

\(\hat{\alpha}\) not unique!
Non-uniqueness of the MLE

• Representational non-uniqueness
  - where to put the masses in the maximal intersections

• Mixture non-uniqueness
  - how to assign masses to the intersections
(Non-)uniqueness of the MLE

• Handling the representational non-uniqueness
  – Lower bound for MLE by assigning all mass to the upper right corners of the maximal intersections
  – Upper bound for MLE by assigning all mass to the lower left corners of the maximal intersections
Sufficient condition for mixture uniqueness

• **Theorem 3.3**

  The MLE $\hat{\alpha}$ is unique if the clique matrix $C$ has rank $m$.

Basic Characterization of the MLE

CURRENT STATUS DATA
Problem Setting

Recall
1. $X$: failure time
   - $X \sim F$
2. $T$: observation time
   - $T \sim G$
3. $X$ independent of $T$
4. $n$ observations: iid copies of

$$(T, \Delta) \equiv (T, 1_{\{X \leq T\}})$$

Goal: Estimate distribution function $F(x) = P[X \leq x]$
Problem Setting

• Notation
  - \( T_{(1)}, \ldots, T_{(n)} \) : order statistics of \( T_1, \ldots, T_n \)
  - \( \Delta_{(1)}, \ldots, \Delta_{(n)} \) : corresponding \( \Delta \) values,

i.e. \( \Delta_{(i)} = \Delta_j \) if \( T_{(i)} = T_j \).

• Recall
  - \( L_n(F) = \prod_{i=1}^{n} F(T_i)^{\Delta_i} (1 - F(T_i))^{1-\Delta_i} \)
  - \( l_n(F) = \sum_{i=1}^{n} \Delta_i \log F(T_i) + (1 - \Delta_i) \log(1 - F(T_i)) \)
Basic Characterization

• Let \( Y = \{ y \in \mathbb{R}^n : 0 < y_1 \leq \ldots \leq y_n < 1 \} \)

• Define \( \hat{y} \) by setting \( \hat{y}_i \equiv \hat{F}_n \left( T_{(i)} \right) \).

an estimator of \( F \)
Basic Characterization

• Corollary 1.2

The vector \( \hat{y} \) is the MLE iff

\[
\forall j \in \{1, \ldots, n\} : \quad \sum_{i<j} \{\Delta_{(i)} - \hat{y}_i\} \geq 0.
\]

If \( \hat{y}_j > \hat{y}_{j-1} \) (with \( \hat{y}_0 = 0, \hat{y}_{n+1} = 1 \)):

\[
\sum_{i<j} \{\Delta_{(i)} - \hat{y}_i\} = 0
\]
Basic Characterization

• Proposition 1.3

- \[ \mathcal{P} = \left\{ P_i = \left( i, \sum_{j \leq i} \Delta_{(j)} \right), i = 0, \ldots, n \right\} \]
- \( H \) greatest convex minorant of \( \mathcal{P} \).

\[ \hat{y} \text{ is MLE iff } \forall i = 1, \ldots n : \]
\[ \hat{y}_i \text{ equals the left derivative of } H \text{ at } i. \]

• Definition

\( H \) is called greatest convex minorant of \( V \), if
- \( H \) is a convex function such that \( H(t) \leq V(t), \forall t, \)
- \( H(t) = V(t) \) holds if \( H \) has a change of slope.
Asymptotic Theory

- **Consistency**: «Does the MLE $\hat{F}_n$ converge to the actual distribution $F_0$?»

- **Rate of convergence**: «How fast does $\hat{F}_n$ converge?»

- **Limiting distribution**: «What does $n^{1/3} (\hat{F}_n(t) - F_0(t))$ converge to?»
Global and local consistency of our estimator

CONSISTENCY
Consistency

«Does the MLE $\hat{F}_n$ converge to the actual distribution $F_0$?»

Definition

An estimator $T_n$ of parameter $\theta$ is said to be consistent if

$$T_n \xrightarrow{P} \theta$$
Consistency

• Global Consistency

• Local Consistency
Global Consistency

For current status data one can show $L_1(G)$ consistency:

$$\int |\hat{F}_n(t) - F_0(t)| dG(t) \xrightarrow{a.s.} 0$$

- Why global?
- Role of $G$?
Local Consistency

From the global consistency one can derive:

Proposition

– $F_0$ continuous at $t_0$

– $G$ continuously differentiable at $t_0$ with strictly positive derivative $g(t_0)$

Then we can choose $r > 0$ such that:

$$\sup_{t \in [t_0-r,t_0+r]} |\hat{F}_n(t) - F_0(t)| \overset{a.s.}{\rightarrow} 0$$
Consistency

«Does the MLE $\hat{F}_n$ converge to the actual distribution $F_0$?»

→ «Yes, $\hat{F}_n$ converges to $F_0$ globally and locally»
Global and local rate of convergence of our estimator

RATE OF CONVERGENCE
Rate Of Convergence

«How fast does $\hat{F}_n$ converge?»
Rate Of Convergence

- Global rate of convergence
- Local rate of convergence
Global Rate Of Convergence

For current status data one can show:

\[ n^{1/3} \int |\hat{F}_n(t) - F_0(t)|dG(t) = O_p(1) \]

→ global rate is \( n^{1/3} \)!

Definition

A sequence of r.v. \( X_1, X_2, \ldots \), is said to be of order \( O_p(1) \) if:

\[ \forall \, \varepsilon > 0 \, \exists \, c, N \text{ such that} \]

\[ P(|X_n| > c) < \varepsilon \text{ for all } n > N \]
Local Rate Of Convergence

**Theorem**

- $0 < F_0(t_0) < 1$
- $F_0$ and $G$ continuously differentiable at $t_0$ with positive derivatives $f_0(t_0)$ resp. $g(t_0)$

Then

$$n^{1/3} \left| \hat{F}_n(t_0) - F_0(t_0) \right| = O_p(1)$$

→ local rate is $n^{1/3}$!
Rate Of Convergence

«How fast does $\tilde{F}_n$ converge?»

→ «It converges at rate $n^{1/3}$, globally and locally»
The limiting distribution of the MLE

LOCAL LIMITING DISTRIBUTION
Local Limiting Distribution

«What does \( n^{1/3} \left( \hat{F}_n(t) - F_0(t) \right) \) converge to?»
Local Limiting Distribution

Definition

- \( W := \text{two sided BM with mean 0 and variance} \)
  \[ E[W(s)W(t)] = (|s| \land |t|)1\{st > 0\}F_0(t_0)(1 - F_0(t_0))/g(t_0) \]
- \( V(t) := W(t) + \frac{1}{2}f_0(t_0)t^2 \)
- \( H(t) := \text{greatest convex minorant of} \ V(t) \)
(see blackboard)

Definition

A **Brownian Motion** \((W_t)\) is a stochastic process with:

- \( W_0 = 0 \)
- \( W_t \text{ continuous a.s.} \)
- \( W_t \text{ has independent increments with} \ W_t - W_s \sim \mathcal{N}(0, t - s) \)
Local Limiting Distribution

**Theorem**

- $0 < F_0(t_0) < 1$
- $F_0$ and $G$ continuously differentiable at $t_0$ with continuous derivatives $f_0(t_0)$ resp. $g(t_0)$

Then

$$n^{1/3} (\hat{F}_n(t_0) - F_0(t_0)) \xrightarrow{d} H'(0)$$
Local Limiting Distribution

«What does $n^{1/3} \left( \hat{F}_n(t) - F_0(t) \right)$ converge to?»

→ «To the slope of the convex minorant of a Brownian motion process plus a parabola»
Likelihood ratio test

- LR-test $\sim \bar{D} = \int (S(t)^2 - S_0(t)^2)dt$.

  - $S$: slope process of greatest convex minorant of a two-sided Brownian motion plus a parabola
  - $S_0$: slope process of greatest convex minorant of a two-sided Brownian motion plus a parabola under the constraints:
    - slopes $\geq 0$, for $t > 0$,
    - Slopes $\leq 0$, for $t < 0$. 
Likelihood ratio test

- Remarks
  - Not a $\chi^2$ – distribution
  - Depends on the slope process of a Brownian motion
  - Independent of the parameters of the problem
Pointwise confidence intervals

• LR-test denoted by $\lambda_n(\theta)$ testing
  - $H_0 : F(t_0) = \theta$
  - $H_1 : F(t_0) \neq \theta$.

• For $0 < \alpha < 1$: let $d_\alpha$ s.t. $P(D > d_\alpha) = \alpha$

• Approximate $1 - \alpha$ confidence set:
  $$C_{n,\alpha} \equiv \{ \theta : 2\log \lambda_n(\theta) \leq d_\alpha \}$$

• It can be shown: $C_{n,\alpha}$ are closed intervals in $(0,1)$ if $X < t_0$ and $T > t_0$. 
Pointwise confidence intervals

• Proposition 4.1

\( F \) and \( G \) have densities \( f \) and \( g \), both positive, continuous in a neighborhood of \( t_0 \). Then, for \( n \rightarrow \infty \):

\[
P_{F,G} \left( F(t_0) \in C_{n,\alpha} \right) \rightarrow P \left( D \leq d_\alpha \right) = 1 - \alpha.
\]
Thank you for your attention!