An Extension of The Koziol-Green Model under Dependent Censoring

Auguste Gaddah
and
Roel Braekers

Universiteit Hasselt
I-BioStat

September 18, 2008
1 Introduction
   • Background
   • Problems
   • Some solutions

2 The Extended Koziol-Green (EKG) Model
   • The EKG estimator
   • Asymptotic results
   • A simulation study
   • Real data application

3 Conclusion

4 References
Background

Time until certain event is often of interest

- time to recurrence of cancer tumor
- length of stay in an environment
- duration of a strike
Some censoring types

Event time of interest (survival time) is mostly not fully observed (Censoring)

- Only known to be smaller than some value (Left Censoring)
- Only known to be between some interval (Interval Censoring)
- Only known to be larger than some value (Right Censoring)
Random Right Censoring

We do not observe

- the true event time $Y_i \sim F$
- the true censoring time $C_i \sim G$

We only observe

- $Z_i = \min(Y_i, C_i) \sim H$
- $\delta_i = I(Y_i \leq C_i)$
- some possible covariates
**Independent and Non-informative Censoring**

- $Y_i$ and $C_i$ are independent
- $F$ and $G$ are not linked

\[ \Downarrow \]

**Kaplan-Meier estimator**

\[
\bar{F}_{KM}^n(t) = \prod_{i: Z_i \leq t} \left( \frac{n - R_i}{n - R_i + 1} \right)^{\delta_i}
\]

where,

- $R_i$ is the rank of $i$th observation in the $Z$ sample
Dependent and Non-informative Censoring

- $Y_i$ and $C_i$ are dependence
  - $S(t_1, t_2) = \varphi^{-1}(\varphi(\bar{F}(t_1)) + \varphi(\bar{G}(t_2)))$
- $F$ and $G$ are not linked

⇓

Copula Graphics (Rivest and Well, 2001)

\[
\bar{F}_{nRW}^R(t) = \varphi^{-1} \left[ \sum_{Z_i \leq t} \left( \varphi(\bar{H}_n(Z_i^-)) + \varphi(\bar{H}_n(Z_i^-) - \frac{1}{n}) \right) \right]
\]

where,

- $H_n$ is the empirical distribution of $Z$
Independent and Informative Censoring

- $Y_i$ and $C_i$ are independence
- $F$ and $G$ are linked
  - $\bar{G}(t) = \bar{F}(t)^{\beta}$, $\beta > 0$

\[ \downarrow \]

Koziol-Green estimator

\[ \bar{F}_{n}^{KG}(t) = \bar{H}_{n}(t)^{\gamma} \quad , \quad \gamma = \frac{1}{1 + \beta} \]

Implicitly assumes $Z_i \perp \delta_i$
Dependent and informative Censoring

- $Y_i$ and $C_i$ are dependent
- $F$ and $G$ are linked
- $Z$ and $\delta$ are independent

\[ \Downarrow \]

Copula Graphics (Braekers and Veraverbeke, 2007)

\[ F_{BV}^n(t) = \varphi^{-1} \left( \gamma_n \varphi(\tilde{H}_n(t)) \right) \]
Dependent and informative Censoring

- $Y_i$ and $C_i$ are dependent
- $F$ and $G$ are linked
- $Z$ and $\delta$ are independent

\[ \Downarrow \]

Copula Graphics (Braekers and Veraverbeke, 2007)

\[ F_{n}^{BV}(t) = \varphi^{-1}(\gamma_n \varphi(\bar{H}_n(t))) \]

What if $Z$ and $\delta$ are not independent?
The EKG estimator

- $Y_i$ and $C_i$ are dependent
- $F$ and $G$ are linked
- $Z_i$ and $\delta_i$ are dependent

$\Downarrow$

**Extended Koziol-Green estimator**

$$
\bar{F}_n^{EKG}(t) = \varphi^{-1} \left( - \int_0^{H_n(t)} \varphi'(1 - w) C^2(\gamma_n, w) \, dw \right)
$$

- $\gamma_n$ is the proportion of uncensored observations
- $\varphi$ is a copula generator function
- $C$ is a copula function
The Extended Koziol-Green (EKG) Model
Asymptotic results

Exponential bound and strong consistency

Under suitable regularity conditions on the copula function $C$, there exist some $M_1, M_2 > 0$ such that

**Theorem 1**

(a) $P \left( \sup_{0 \leq t \leq T} |F_n(t) - F(t)| > \varepsilon \right) \leq 2 \exp \left( - \frac{n\varepsilon^2}{4M_1(\varepsilon + 2M_1\gamma)} \right) + 2 \exp \left( - \frac{n\varepsilon^2}{2M_2^2} \right)$

(b) If $\frac{\log(n)}{n} \to 0$ as $n \to \infty$, then

$\sup_{0 \leq t \leq T} |F_n(t) - F(t)| \to 0 \quad a.s.$
Almost sure representation and weak convergence

Theorem 2

\((a)\) \(F_n(t) - F(t) = \frac{1}{n} \sum_{i=1}^{n} k(t; Z_i, \delta_i) + r_n(t)\)

and as \(n \to \infty\),

\[
\sup_{0 \leq t \leq T} |r_n(t)| = O(n^{-1} \log(n)) \quad \text{a.s.}
\]

where

- \(k(t; Z_i, \delta_i) = \frac{1}{\varphi'(F(t))} \left[ (1\{\delta_i = 1\} - \gamma) \int_{o}^{H(t)} \varphi'(1 - w) C^2(\gamma_n, w) \, dw \right.\)

\[
+ \left. (1\{Z_i \leq t\} - H(t)) \varphi'(\bar{H}(t)) C^2(\gamma_n, H_n(t)) \right]
\]
The Extended Koziol-Green (EKG) Model

Almost sure representation and weak convergence

Theorem 2

(b) Suppose \( \frac{\log(n)}{n} \to 0 \) as \( n \to \infty \), then

\[
\sqrt{n} \left( F_n(\cdot) - F(\cdot) \right) \to W(\cdot)
\]

With covariance function

\[
\Gamma(s, t) = \frac{1}{\varphi'(F(s))\varphi'(F(t))} \left[ \gamma(1 - \gamma) \int_0^{H(s)} \varphi' (1 - w) C^{12} (\gamma, w) \, dw \int_0^{H(t)} \varphi' (1 - w) C^{12} (\gamma, w) \, dw \\
+ (H^u(s) - \gamma H(s)) \varphi' (\bar{H}(s)) C^2 (\gamma, H(s)) \int_0^{H(t)} \varphi' (1 - w) C^{12} (\gamma, w) \, dw \\
+ (H^u(t) - \gamma H(t)) \varphi' (\bar{H}(t)) C^2 (\gamma, H(t)) \int_0^{H(s)} \varphi' (1 - w) C^{12} (\gamma, w) \, dw \\
+ (H(s \wedge t) - H(s)H(t)) \varphi' (\bar{H}(s))\varphi' (\bar{H}(t)) C^2 (\gamma, H(s)) C^2 (\gamma, H(t)) \right]
\]
Data generation

1. we generate two independent uniform \((0,1)\) random variables \(u\) and \(t\).

2. we set \(v = c_u^{-1}(t)\), where
   \[
   c_u(v) = \frac{\partial}{\partial u} \left\{ \varphi_{-1}^{-1}(\varphi_0(u) + \varphi_0(v)) \right\}
   \]

3. we solve for \(h\) in
   \[
   1 - h = \varphi^{-1} \left( -\int_0^h \varphi'(1-w)C^2(\gamma, w) \, dw + \varphi(u) \right)
   \]

4. we set
   \[
   \begin{align*}
   Y &= -\frac{1}{\lambda} \log(v) \\
   C &= -\frac{1}{\lambda} \log \left( \varphi^{-1}(\varphi(1-h) - \varphi(u)) \right)
   \end{align*}
   \]

5. we set \(Z_i = \min(Y_i, C_i)\) and \(\delta_i = 1\{Y_i \leq C_i\}\)
Assuming the Clayton copula for $Y$ and $C$; and Clayton copula for $Z$ and $\delta$

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>$N$</th>
<th>$\gamma = 0.25$</th>
<th>$\gamma = 0.50$</th>
<th>$\gamma = 0.75$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-0.6$</td>
<td>50</td>
<td>0.1772</td>
<td>0.0302</td>
<td>0.0204</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.1342</td>
<td>0.0159</td>
<td>0.0107</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>0.1239</td>
<td>0.0102</td>
<td>0.0075</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.1179</td>
<td>0.0090</td>
<td>0.0045</td>
</tr>
<tr>
<td>$0.6$</td>
<td>50</td>
<td>0.1502</td>
<td>0.0154</td>
<td>0.0123</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.1223</td>
<td>0.0077</td>
<td>0.0064</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>0.1109</td>
<td>0.0046</td>
<td>0.0037</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.1047</td>
<td>0.0039</td>
<td>0.0034</td>
</tr>
</tbody>
</table>
Assuming the Clayton copula for $Y$ and $C$; and Frank copula for $Z$ and $\delta$

<table>
<thead>
<tr>
<th>N</th>
<th>$\gamma = 0.25$</th>
<th>$\gamma = 0.50$</th>
<th>$\gamma = 0.75$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.1511</td>
<td>0.0314</td>
<td>0.0156</td>
</tr>
<tr>
<td>100</td>
<td>0.1268</td>
<td>0.0164</td>
<td>0.0036</td>
</tr>
<tr>
<td>150</td>
<td>0.1174</td>
<td>0.0105</td>
<td>0.0030</td>
</tr>
<tr>
<td>200</td>
<td>0.1132</td>
<td>0.0073</td>
<td>0.0022</td>
</tr>
<tr>
<td>50</td>
<td>0.1782</td>
<td>0.0131</td>
<td>0.0128</td>
</tr>
<tr>
<td>100</td>
<td>0.1448</td>
<td>0.0072</td>
<td>0.0062</td>
</tr>
<tr>
<td>150</td>
<td>0.1225</td>
<td>0.0043</td>
<td>0.0043</td>
</tr>
<tr>
<td>200</td>
<td>0.1057</td>
<td>0.0034</td>
<td>0.0033</td>
</tr>
</tbody>
</table>
Dataset: Survival with malignant melanoma (skin cancer)

- Study on survival rate of malignant melanoma patients
- Event time
  - time elapsed from surgery until death from malignant melanoma
- Censoring time
  - end of study
  - death from other diseases
- Total of 205 malignant melanoma patients
- 57 (28%) events and 148 (72%) censored
- Weaker patients are more likely to die
Independence assumption

To check independence of observed time and censoring indicator

Relation

\[
P(\delta = 1|Z \leq t) = \frac{P(\delta = 1, Z \leq t)}{P(Z \leq t)} = \frac{P(\delta = 1)P(Z \leq t)}{P(Z \leq t)} \quad \text{if} \quad Z \perp \delta
\]

- The \( \frac{H^u_n(t)}{H_n(t)} \) is constant over time iff \( Z \perp \delta \)
The Extended Koziol-Green (EKG) Model

Real data application

Independence assumption
Clayton copula for the latent variables
Frank copula for the latent variables
- Clayton copula for latent variable = Left Panel
- Frank copula for latent variable = Right Panel
Conclusion

- Common problem in time to event analysis is censoring

- Need for dependence structure between unobserved variables (event and censoring time)

- Need for dependence structure between observed time and censoring indicator

- Inference is influenced by assumed dependence structures


