On a Measure of Information Gain for Regression Models in Survival Analysis

Pascal Roy
Delphine Maucort-Boulch, Janez Stare

Equipe Biostatistique Santé, UMR CNRS 5558

26th May 2010
Outline

1. Information gain measure
   - Explained variation
   - Expected likelihood ratio

2. Simulations study
   - Method
   - Results

3. Conclusion
In linear regression

\[ \sum_{i} (y_i - \bar{y})^2 = \sum_{i} (\hat{y}_i - \bar{y})^2 + \sum_{i} (y_i - \hat{y}_i)^2 \]

\[ SS_{tot} = SS_{reg} + SS_{res} \]

\[ \rho^2 = \frac{\sum_{i} (\hat{y}_i - \bar{y})^2}{\sum_{i} (y_i - \bar{y})^2} = \frac{SS_{reg}}{SS_{tot}} \]
In survival models

The most comprehensive (Kent et al)

\[ \rho_{IG}^2 = 1 - e^{-E(LR)} \]

Softwares output

\[ \hat{\rho}_n^2 = 1 - e^{-\frac{2}{n} \left[ \text{loglik}_{\text{model}} - \text{loglik}_{\text{null}} \right]} \]

But no statistical justification

O’Quigley et al proposal

\[ \hat{\rho}_{IG}^2 = 1 - e^{-\frac{2}{\#\text{events}} \left[ \text{loglik}_{\text{model}} - \text{loglik}_{\text{null}} \right]} \]
Population value

We are interested in

\[ E(LR) = 2 \int_0^\infty \log \left( \frac{f_M(t)}{f_0(t)} \right) dF_M(t) \]

Twice the Kullback Leibler information gain

The suggestion by O’Quigley et al was reported to be biased under censoring, although less than the \( \rho_n^2 \)

- The jumps in the survival curve are not all equal to \( 1/k \)
- The overall survival curve may not drop to 0
Last observed failure time $\tau$

When the last observed failure occurs at time $\tau$ because of censoring or because we want to limit ourselves to observations less than a given time $\tau$

$$E(LR) = 2 \int_0^\infty \log \frac{f_M(t)}{f_0(t)} dF_M(t|\tau) = 2 \int_0^\infty \log \frac{f_M(t)}{f_0(t)} \frac{dF_M(t)}{F_M(\tau)}$$
Two types of censoring

Before the last failure time $\tau$

Attenuation of the sample size and, if random, only affect the variability of the estimator, so efficiency, not its expected value → Weights, as jumps in the survival curve, should compensate for the missing information

$$\hat{E}_w(LR) = 2 \sum_{1}^{k} \log(LR) \frac{\Delta \hat{F}_M(t)}{\hat{F}_M(\tau)}$$

After the last failure time $\tau$

No information on precision after $\tau$ and since a measure of predictive accuracy is an overall measure, it will be affected by such censoring → impute under the model
Data generation

**Complete data**
- Covariate
  - continuous \( U[0, \sqrt{3}] \)
  - binary
- \( \beta \in \{1, 2, 5\} \)
- Times generated under exponential model
- \( n \in \{200, 500, 1000, 5000\} \)
- Cox model fitted
- 100 iterations

**Censoring**
1. Complete data censored in two different ways
   1. random censoring
      \( \tau \) highest failure time determined
   2. Type 1 censoring
      all times greater than \( \tau \) are censored
2. Uniform censoring, percentages from 10% to 90%
## Data generation

### Complete data
- **Covariate**
  - continuous $U[0, \sqrt{3}]$
  - binary
- $\beta \in \{1, 2, 5\}$
- Times generated under exponential model
- $n \in \{200, 500, 1000, 5000\}$
- Cox model fitted
- 100 iterations

### Censoring
1. Complete data censored in two different ways
   - 1. random censoring $\tau$ highest failure time determined
   - 2. Type 1 censoring all times greater than $\tau$ are censored
2. Uniform censoring, percentages from 10% to 90%
### Measures

#### Different estimations

<table>
<thead>
<tr>
<th>Estimation</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_n^2$</td>
<td>$1 - \exp\left(\frac{2}{n} \sum_{1}^{k} \log(\hat{L}R)\right)$</td>
</tr>
<tr>
<td>$\rho_k^2$</td>
<td>$1 - \exp\left(\frac{2}{k} \sum_{1}^{k} \log(\hat{L}R)\right)$</td>
</tr>
<tr>
<td>$\rho_w^2$</td>
<td>$1 - \exp\left(2 \sum_{1}^{k} \log(\hat{L}R) \frac{\Delta \hat{F}_M(t)}{\hat{F}_M(t)}\right)$</td>
</tr>
<tr>
<td>$\rho_i^2$</td>
<td>$1 - \exp\left(2 \sum_{1}^{k'} \log(\hat{L}R) \Delta \hat{F}_M(t)\right)$</td>
</tr>
</tbody>
</table>

#### Three data sets

1. Randomly censored
2. Censored after $\tau$
3. Complete

$\rho_w^2$: Kaplan-Meier estimates of $\hat{F}_M(t)$

$\rho_i^2$: average on 10 imputations
log likelihood ratio components over time

Times
$\rho^2$ over time
Randomly censored data

Population value 0.15 and 0.38 for continuous variable with
\( \beta \in \{1, 2\} \) respectively

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>Size</th>
<th>%</th>
<th>( \rho_n )</th>
<th>se</th>
<th>( \rho_k )</th>
<th>se</th>
<th>( \rho_w )</th>
<th>se</th>
<th>( \rho_i )</th>
<th>se</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5000</td>
<td>80</td>
<td>0.04</td>
<td>0.01</td>
<td>0.19</td>
<td>0.03</td>
<td>0.19</td>
<td>0.04</td>
<td>0.17</td>
<td>0.03</td>
</tr>
<tr>
<td>1</td>
<td>5000</td>
<td>50</td>
<td>0.09</td>
<td>0.01</td>
<td>0.18</td>
<td>0.01</td>
<td>0.17</td>
<td>0.02</td>
<td>0.15</td>
<td>0.02</td>
</tr>
<tr>
<td>1</td>
<td>5000</td>
<td>20</td>
<td>0.13</td>
<td>0.01</td>
<td>0.16</td>
<td>0.01</td>
<td>0.15</td>
<td>0.01</td>
<td>0.14</td>
<td>0.01</td>
</tr>
<tr>
<td>1</td>
<td>200</td>
<td>80</td>
<td>0.04</td>
<td>0.03</td>
<td>0.20</td>
<td>0.12</td>
<td>0.21</td>
<td>0.15</td>
<td>0.16</td>
<td>0.11</td>
</tr>
<tr>
<td>1</td>
<td>200</td>
<td>50</td>
<td>0.10</td>
<td>0.03</td>
<td>0.19</td>
<td>0.06</td>
<td>0.18</td>
<td>0.07</td>
<td>0.15</td>
<td>0.06</td>
</tr>
<tr>
<td>1</td>
<td>200</td>
<td>20</td>
<td>0.12</td>
<td>0.04</td>
<td>0.15</td>
<td>0.05</td>
<td>0.14</td>
<td>0.05</td>
<td>0.14</td>
<td>0.05</td>
</tr>
<tr>
<td>2</td>
<td>5000</td>
<td>80</td>
<td>0.13</td>
<td>0.01</td>
<td>0.50</td>
<td>0.02</td>
<td>0.51</td>
<td>0.04</td>
<td>0.42</td>
<td>0.03</td>
</tr>
<tr>
<td>2</td>
<td>5000</td>
<td>50</td>
<td>0.29</td>
<td>0.01</td>
<td>0.50</td>
<td>0.01</td>
<td>0.48</td>
<td>0.02</td>
<td>0.39</td>
<td>0.02</td>
</tr>
<tr>
<td>2</td>
<td>5000</td>
<td>20</td>
<td>0.37</td>
<td>0.01</td>
<td>0.44</td>
<td>0.01</td>
<td>0.40</td>
<td>0.01</td>
<td>0.39</td>
<td>0.01</td>
</tr>
<tr>
<td>2</td>
<td>200</td>
<td>80</td>
<td>0.13</td>
<td>0.04</td>
<td>0.49</td>
<td>0.12</td>
<td>0.47</td>
<td>0.17</td>
<td>0.38</td>
<td>0.12</td>
</tr>
<tr>
<td>2</td>
<td>200</td>
<td>50</td>
<td>0.29</td>
<td>0.05</td>
<td>0.49</td>
<td>0.08</td>
<td>0.46</td>
<td>0.09</td>
<td>0.38</td>
<td>0.07</td>
</tr>
<tr>
<td>2</td>
<td>200</td>
<td>20</td>
<td>0.36</td>
<td>0.05</td>
<td>0.43</td>
<td>0.05</td>
<td>0.40</td>
<td>0.05</td>
<td>0.38</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Uncensored before $\tau$ data

Population value 0.15 and 0.38 for continuous variable with $\beta \in \{1, 2\}$ respectively

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>Size</th>
<th>%</th>
<th>$\rho_n$</th>
<th>se</th>
<th>$\rho_w$</th>
<th>se</th>
<th>$\rho_i$</th>
<th>se</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5000</td>
<td>80</td>
<td>0.07</td>
<td>0.01</td>
<td>0.19</td>
<td>0.02</td>
<td>0.17</td>
<td>0.02</td>
</tr>
<tr>
<td>1</td>
<td>5000</td>
<td>50</td>
<td>0.13</td>
<td>0.01</td>
<td>0.17</td>
<td>0.01</td>
<td>0.15</td>
<td>0.01</td>
</tr>
<tr>
<td>1</td>
<td>5000</td>
<td>20</td>
<td>0.14</td>
<td>0.01</td>
<td>0.15</td>
<td>0.01</td>
<td>0.15</td>
<td>0.01</td>
</tr>
<tr>
<td>1</td>
<td>200</td>
<td>80</td>
<td>0.07</td>
<td>0.03</td>
<td>0.21</td>
<td>0.09</td>
<td>0.18</td>
<td>0.08</td>
</tr>
<tr>
<td>1</td>
<td>200</td>
<td>50</td>
<td>0.13</td>
<td>0.04</td>
<td>0.18</td>
<td>0.05</td>
<td>0.16</td>
<td>0.05</td>
</tr>
<tr>
<td>1</td>
<td>200</td>
<td>20</td>
<td>0.14</td>
<td>0.04</td>
<td>0.15</td>
<td>0.05</td>
<td>0.14</td>
<td>0.04</td>
</tr>
<tr>
<td>2</td>
<td>5000</td>
<td>80</td>
<td>0.21</td>
<td>0.01</td>
<td>0.51</td>
<td>0.02</td>
<td>0.42</td>
<td>0.01</td>
</tr>
<tr>
<td>2</td>
<td>5000</td>
<td>50</td>
<td>0.37</td>
<td>0.01</td>
<td>0.48</td>
<td>0.01</td>
<td>0.39</td>
<td>0.01</td>
</tr>
<tr>
<td>2</td>
<td>5000</td>
<td>20</td>
<td>0.39</td>
<td>0.01</td>
<td>0.40</td>
<td>0.01</td>
<td>0.39</td>
<td>0.01</td>
</tr>
<tr>
<td>2</td>
<td>200</td>
<td>80</td>
<td>0.19</td>
<td>0.05</td>
<td>0.49</td>
<td>0.09</td>
<td>0.41</td>
<td>0.07</td>
</tr>
<tr>
<td>2</td>
<td>200</td>
<td>50</td>
<td>0.35</td>
<td>0.06</td>
<td>0.47</td>
<td>0.08</td>
<td>0.38</td>
<td>0.05</td>
</tr>
<tr>
<td>2</td>
<td>200</td>
<td>20</td>
<td>0.38</td>
<td>0.05</td>
<td>0.40</td>
<td>0.05</td>
<td>0.38</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Weights and imputation correct for the bias of O’Quigley et al proposal
Gain in estimation comes with a price - a bigger variance
Two groups case

Fit with exponential model: \( \beta = 5.01, \rho_{IG}^2 = 0.975 \)
Fit with Weibull model: \( \beta = 5.01, \rho_{IG}^2 = \)
Fit with Cox model: \( \beta = 5.03, \rho_{IG}^2 = \rho_{IG}^2 = \)

Some measures: \( \beta \to \infty \to 1 \)
exponential distribution
\( \beta = 5 \)
sample size 1000

R output
\( \text{Rsquare}= 0.711 \)
(max possible = 1)
Two groups case

Fit with exponential model: \( \beta = 5.01, \rho^2_{IG} = 0.975 \)
Fit with Weibull model: \( \beta = 5.01, \rho^2_{IG} = \)
Fit with Cox model: \( \beta = 5.03, \rho^2_{IG} = \)

some measures \( \beta \to \infty \to 1 \)

exponential distribution
\( \beta = 5 \)
sample size 1000

R output
Rsquare= 0.711
(max possible=1)
Two groups case

Fit with exponential model: $\beta = 5.01$, $\rho_{IG}^2 = 0.975$

Fit with Weibull model: $\beta = 5.01$, $\rho_{IG}^2 = \ldots$  

Fit with Cox model: $\beta = 5.03$, $\rho_{IG}^2 = \ldots$  

R output:
Rsquare= 0.711
(max possible=1)

some measures $\beta \to \infty$ \Rightarrow 1

exponential distribution
$\beta = 5$

sample size 1000

Kent O'Quigley
Two groups case

Fit with exponential model: \( \beta = 5.01, \rho_{IG}^2 = 0.975 \)
Fit with Weibull model: \( \beta = 5.01, \rho_{IG}^2 = 0.830 \)
Fit with Cox model: \( \beta = 5.03, \rho_{IG}^2 = 0.701 \)

some measures \( \beta \rightarrow \infty \rightarrow 1 \)

exponential distribution
\( \beta = 5 \)
sample size 1000

R output
Rsquare= 0.711
(max possible=1)
Two groups case

- Fit with exponential model: $\beta = 5.01, \rho_{IG}^2 = 0.975$
- Fit with Weibull model: $\beta = 5.01, \rho_{IG}^2 = 0.830$
- Fit with Cox model: $\beta = 5.03, \rho_{IG}^2 = 0.701$

Some measures $\beta \to \infty$ 1

Exponential distribution
$\beta = 5$
Sample size 1000

R output
Rsquare= 0.711
(max possible=1)
Two groups case

some measures $\beta \to \infty$ 1

exponential distribution
$\beta = 5$
sample size 1000

Fit with exponential model: $\beta = 5.01$, $\rho_{IG}^2 = 0.975$
Fit with Weibull model: $\beta = 5.01$, $\rho_{IG}^2 = 0.830$
Fit with Cox model: $\beta = 5.03$, $\rho_{IG}^2 = 0.711$

Kent O’Quigley

R output
Rsquare= 0.711
(max possible=1)
Two groups case

Fit with exponential model: \( \beta = 5.01, \rho_{IG}^2 = 0.975 \)
Fit with Weibull model: \( \beta = 5.01, \rho_{IG}^2 = 0.830 \)
Fit with Cox model: \( \beta = 5.03, \rho_{IG}^2 = 0.711 \)

some measures \( \beta \rightarrow \infty \rightarrow 1 \)

exponential distribution
\( \beta = 5 \)
sample size 1000

R output
Rsquare= 0.711
(max possible=1)
Two groups case

Fit with exponential model: $\beta = 5.01$, $\rho_{IG}^2 = 0.975$

Fit with Weibull model: $\beta = 5.01$, $\rho_{IG}^2 = 0.830$

Fit with Cox model: $\beta = 5.03$, $\rho_{IG}^2 = 0.711$

$\rho_{IG}^2 = 0.832$

some measures $\beta \to \infty \to 1$

exponential distribution
$\beta = 5$

sample size 1000

R output
Rsquare= 0.711
(max possible=1)
Two groups case

Fit with exponential model: $\beta = 5.01$, $\rho_{IG}^2 = 0.975$
Fit with Weibull model: $\beta = 5.01$, $\rho_{IG}^2 = 0.830$
Fit with Cox model: $\beta = 5.03$, $\rho_{IG}^2 = 0.711$
$\rho_{IG}^2 = 0.832$

Kent O’Quigley
Two groups case

Fit with exponential model: \( \beta = 5.01, \rho_{IG}^2 = 0.975 \)
Fit with Weibull model: \( \beta = 5.01, \rho_{IG}^2 = 0.830 \)
Fit with Cox model: \( \beta = 5.03, \rho_{IG}^2 = 0.711, \rho_{IG} = 0.832 \)

some measures \( \beta \to \infty \) \( \to 1 \)

exponential distribution
\( \beta = 5 \)
sample size 1000

R output
Rsquare = 0.711
(max possible = 1)
Two groups case

Fit with exponential model: \( \beta = 5.01, \rho^{2}_{IG} = 0.975 \)
Fit with Weibull model: \( \beta = 5.01, \rho^{2}_{IG} = 0.830 \)
Fit with Cox model: \( \beta = 5.03, \rho^{2}_{IG} = 0.711 \)

\( \rho^{2}_{IG} = 0.832 \)

R output

Rsquare= 0.711
(max possible=1)
Two groups case

Fit with exponential model: \( \beta = 5.01, \rho_{IG}^2 = 0.975 \)
Fit with Weibull model: \( \beta = 5.01, \rho_{IG}^2 = 0.830 \)
Fit with Cox model: \( \beta = 5.03, \rho_{IG}^2 = 0.711 \)
\( \rho_{IG}^2 = 0.832 \)

some measures \( \beta \to \infty \Rightarrow 1 \)

exponential distribution
\( \beta = 5 \)
sample size 1000

R output
\( \text{Rsquare}= 0.711 \)
(max possible=1)
Extensions to parametric models

- Log likelihood

\[
\sum_{\text{uncensored}} \ln f(t|\beta) + \sum_{\text{censored}} \ln S(t|\beta)
\]

- Influence of time values
References