The Bootstrap Method

and some

Reserving applications

"Good simple ideas, ..., are our most precious intellectual commodity, so there is no need to apologize for the easy mathematical level."

Bradley Efron

Presented at GISG 1995 by Christian Larsen
The Bootstrap method was introduced by Bradley Efron, Stanford University in 1979. Efron (1982) begins with some general statements, such as:

"Good simple ideas,..., are our most precious intellectual commodity, so there is no need to apologize for the easy mathematical level."

and

"An important theme of what follows, is the substitution of computational power for theoretical analysis. The payoff, of course, is the freedom from the constraints of traditional parametric theory, with its overreliance on a small set of standard models for which theoretical solutions are available. In the long run, ....(bootstrap) ..., should make clearer the virtues of parametric theory,..."

and

"From a traditional point of view, .... the methods ... are prodigious computational spendthrifts. We blithely ask the reader to consider techniques which require the usual statistical calculations to be multiplied a thousand times over. None of this would have been feasible twenty-five years ago, before the era of cheap and fast computation."

The situation

We consider a random sample \((X_1, ..., X_N)\) of random size \(N\). The distribution \(G\) of \(N\) is assumed to be known and the random variables are assumed to be independent and equally distributed with unknown distribution \(F\):

\[
(X_1, ..., X_N) = (x_1, ..., x_N), \quad X_r \sim F, \quad r=1, ..., N \quad \text{and} \quad N \sim G
\]

It is further assumed that \(X_r\) is independent of the frequency \(N\).

\(X=(X_1, ..., X_N)\) denotes the random sample and \(x=(x_1, ..., x_N)\) the observed realisation.

The problem

Let \(R(X)\) be a function of \(X\). Then \(R(X)\) is a stochastic variable with a distribution that is dependent on \(G\) and of the unknown \(F\). The problem is to estimate the distribution of \(R\) on the basis of the observation \(x\).
A solution - the Bootstrap Method

If the distribution of $F$ was known then the distribution of $R$ could, in theory, be calculated exactly or in practice be approximated with unlimited accuracy by the Monte Carlo method. The simple idea in Bootstrap is to substitute the distribution of $F$ with the empirical distribution based on the observation $x$:

A. Construct the sample probability distribution $E$ (i.e. the empirical distribution of $X$), putting mass $1/n$ at each point $x_1,...,x_n$.

B. Consider the random sample $Y=(Y_1,...,Y_N)$, $Y_r \sim E$, $r=1,...,N$, $N \sim G$.

C. Approximate the distribution of $R(X)$ by the (Bootstrap) distribution of $R(Y)$.

The essential facts are that

- the distribution of $R(Y)$ is only dependent on $E$ and $G$ and if $E$ is a good approximation to $F$ then one can expect that the distribution of $R(Y)$ will be a good approximation to the distribution of $R(X)$ and
- the distribution of $(E,G)$ is known and therefore, the distribution of $R(Y)$ can be calculated.

Application I: Calculation of the uncertainty of reserving estimates

Let $N$ denote the number of observed claims in a specific period and assume that $N$ is Poisson distributed with a known mean. Let $X_r$, $r=1,...,N$, denote the information linked to the claims including information about accident period, development period and payments. Only information regarding the past is available and for some claims $X_r$, all payments and periods of payments are not necessarily included.

Let $RES$ denote the stochastic total of the future payments, i.e. the stochastic reserve.

Let $R_J(X)$ denote a reserve estimator, e.g. the Chain Ladder estimator based on the claim information. $R_J(X)$ is a function of the claims $X$ and therefore a stochastic variable.

The problem we wish to solve is to estimate the distribution of the Chain Ladder estimator $R_J(X)$ rather than just the Chain Ladder point estimate $R_J(x)$. 

225
Example 1

Consider the accident periods 1985-1994 and assume that the number of claims with accident date in period \( i=1985, \ldots, 1994 \) and notification delay \( j=0, \ldots, 9 \) measured by years from accident period to notification period is Poisson\((A_i, p)\) with the parameters outlined below in figure 1.

Assume further that each claim has only one payment which is Gamma distributed with mean \( B_i=(1.05)^{i+0(1985)} \) dependent on the accident period \( i \) and constant coefficient of variation equal to 2. The waiting time \( k=0, \ldots, 9 \) from year of notification to year of payment is assumed to be independent of the waiting time to notification and with distribution outlined in figure 1 below. The distribution of the waiting time \( r \) from accident year to year of payment, i.e. the convolution of \( p \) and \( q \), is also calculated.

\[
\begin{array}{|c|c|c|}
\hline
\text{Period} & A_j & B_j \\
\hline
1985 & 500 & 1.000 \\
1986 & 500 & 1.050 \\
1987 & 500 & 1.103 \\
1988 & 500 & 1.158 \\
1989 & 500 & 1.216 \\
1990 & 500 & 1.276 \\
1991 & 500 & 1.340 \\
1992 & 500 & 1.407 \\
1993 & 500 & 1.477 \\
1994 & 500 & 1.551 \\
\hline
\text{All} & 5000 & 12.578 \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|c|c|}
\hline
\text{Delay} & j \sim p & k \sim q & j+k \sim r \\
\hline
0 & 0.30 & 0.35 & 0.1050 \\
1 & 0.25 & 0.20 & 0.1475 \\
2 & 0.20 & 0.20 & 0.1800 \\
3 & 0.15 & 0.15 & 0.1875 \\
4 & 0.07 & 0.10 & 0.1620 \\
5 & 0.03 & 0 & 0.1095 \\
6 & 0 & 0 & 0.0625 \\
7 & 0 & 0 & 0.0315 \\
8 & 0 & 0 & 0.0115 \\
9 & 0 & 0 & 0.0030 \\
\hline
\text{Mean} & E(j)=1.53 & E(k)=1.45 & E(j+k)=2.98 \\
\hline
\end{array}
\]

On the basis of these distribution assumptions we get \( E(\text{RES}) = 2142 \) and the expected total claim amount is \( 500 \times 12.578 = 6289 \). As a consequence, the expected amount paid already is therefore 4147.

Two data sets, PAST and FUTURE, are generated by simulation in The SAS® System on the basis of the distribution assumptions above. The PAST file consists of 4699 claims where \( i+j<1995 \) and of 3486 payments where \( i+j+k<1995 \). The Future file consists of the remaining data, i.e. claims with date of notification or date of payment in 1995 or later. The sum of future payments, i.e. the observed value of RES, is 2004, compared to the expected value 2142.
As an example, the claim information regarding three claims from the PAST file is shown below:

<table>
<thead>
<tr>
<th>Observation</th>
<th>Claim identification</th>
<th>Accident period i</th>
<th>Transaction</th>
<th>Delay j or j+k</th>
<th>Payment</th>
</tr>
</thead>
<tbody>
<tr>
<td>x_{108}</td>
<td>108</td>
<td>1985</td>
<td>Notification</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>108</td>
<td>1985</td>
<td>Payment</td>
<td></td>
<td>2251.89</td>
</tr>
<tr>
<td>x_{3086}</td>
<td>3086</td>
<td>1991</td>
<td>Notification</td>
<td>1</td>
<td>1666.06</td>
</tr>
<tr>
<td></td>
<td>3086</td>
<td>1991</td>
<td>Payment</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>x_{4649}</td>
<td>4649</td>
<td>1994</td>
<td>Notification</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

![figure 2]

The PAST data is now triangulated and the Chain Ladder estimate $R_f(x)$ for the reserve is calculated. The results are outlined in figure 3 below.

<table>
<thead>
<tr>
<th>ACC. PERIOD</th>
<th>Development Period</th>
<th>Observed Payments</th>
<th>Total Reserve</th>
<th>Total Estimated Payment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1985</td>
<td>50 81 113 97 87 55 36 4 2 1</td>
<td>525</td>
<td>0</td>
<td>525</td>
</tr>
<tr>
<td>1986</td>
<td>59 64 103 85 95 51 43 11 5 1</td>
<td>516</td>
<td>1</td>
<td>516</td>
</tr>
<tr>
<td>1987</td>
<td>52 85 132 82 78 54 38 12 4 1</td>
<td>534</td>
<td>4</td>
<td>539</td>
</tr>
<tr>
<td>1988</td>
<td>77 66 94 79 127 44 29 9 3 1</td>
<td>515</td>
<td>14</td>
<td>528</td>
</tr>
<tr>
<td>1989</td>
<td>81 111 123 115 87 84 45 12 4 1</td>
<td>602</td>
<td>62</td>
<td>665</td>
</tr>
<tr>
<td>1990</td>
<td>79 116 84 96 118 64 42 11 4 1</td>
<td>494</td>
<td>122</td>
<td>616</td>
</tr>
<tr>
<td>1991</td>
<td>79 67 112 132 109 65 42 11 4 1</td>
<td>389</td>
<td>231</td>
<td>621</td>
</tr>
<tr>
<td>1992</td>
<td>74 121 138 125 128 76 50 13 5 1</td>
<td>333</td>
<td>397</td>
<td>730</td>
</tr>
<tr>
<td>1993</td>
<td>78 123 143 129 132 78 51 13 5 1</td>
<td>200</td>
<td>551</td>
<td>752</td>
</tr>
<tr>
<td>1994</td>
<td>112 148 185 167 171 101 67 17 6 2</td>
<td>112</td>
<td>864</td>
<td>976</td>
</tr>
<tr>
<td>ALL</td>
<td>740 983 1226 1107 1132 672 442 113 42 10</td>
<td>4230</td>
<td>2247</td>
<td>6467</td>
</tr>
</tbody>
</table>

![figure 3]

The estimated run off and the model run off are shown in figure 4. For example it is seen that on average 11.4% of the claim amount is estimated to be paid in the accident year, and that the model proportion $r$ is 10.5%.

Run-off pattern

<table>
<thead>
<tr>
<th></th>
<th>Est.</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.4%</td>
<td>11.2%</td>
<td>11.3%</td>
</tr>
<tr>
<td>10.0%</td>
<td>17.1%</td>
<td>17.5%</td>
</tr>
<tr>
<td>10.4%</td>
<td>6.8%</td>
<td>1.7%</td>
</tr>
<tr>
<td>0.6%</td>
<td>1.2%</td>
<td>1.7%</td>
</tr>
<tr>
<td>0.2%</td>
<td>65.3%</td>
<td>34.7%</td>
</tr>
</tbody>
</table>

![figure 4]
Figure 5 shows the observed and the estimated proportion of the estimated ultimate claim costs paid at different times. It is seen that at the end of 31 December 1994 the amount 4220 (65.3%) was paid and that the amount expected to be paid in the future, i.e. the outstanding claims reserve is 2247 (34.7%).

Development of Accumulated Payments / Reserves

In order to calculate the estimation uncertainty of the Chain Ladder estimate 2247, the Bootstrap method has been applied using Larsen & Partners' Actuarial Claims Reserving System. The estimated distribution of $R(Y)$ (figure 6) is based on 200 Monte Carlo simulations from the empirical distribution. The procedure is:

1. Draw a random number $m$ from $G = \text{Poisson}(4699)$.
2. Draw a random sample, $y_1, \ldots, y_m$ from $E$.
3. Calculate the Chain Ladder reserve $R_f(y_1, \ldots, y_m)$ based on $y_1, \ldots, y_m$.
4. Repeat 1-3 200 times.

228
Normally we do not know the distribution of $F$ and a test of the quality of the method is difficult to define. However, in the situation above, where the distribution of $F$ is known, we can easily calculate the distribution of $R(X)$ by simulation:

1. Draw a random number $m$ from $G = \text{Poisson}(4699)$.
2. Draw a random sample, $x_1, \ldots, x_m$ from $F$.
3. Calculate the Chain Ladder reserve $R_j(x_1, \ldots, x_m)$ based on $x_1, \ldots, x_m$.
4. Repeat 1-3 200 times. The results are shown in figure 7.
The reserve estimate \( R_f(x) \) based on the observation \( x \) is exceeding the expected value \( E(RES) \) by approximately 105 and this error is inherited in the entire distribution of \( R_f(Y) \). However, the Bootstrap error distribution has no systematic error and it approximates to the actual error distribution reasonably well, see figure 8 below.

| Mean   | 2253 | 2130 | 0   | 0   |
| STD    | 214  | 179  | 214 | 179 |
| Median | 2247 | 2118 | -6  | -12 |
| 75% fraction | 2373 | 2229 | 120 | 99  |
| 95% fraction | 2622 | 2456 | 369 | 326 |
| 98% fraction | 2713 | 2508 | 460 | 378 |

Please note, that the distribution of \( R_f(Y) \) and of the error \( R_f(Y) - E(R_f(Y)) \) is calculated without using the knowledge of the underlying distribution \( F \).

**Application 2: Estimation of the total uncertainty**

We consider a reserving method \( R \). The total uncertainty consists of the estimation uncertainty (related to the past) plus the uncertainty related to the future payments \( RES - E(RES) \). We assume that the selected model is 'correct' so that the reserve estimate is unbiased, i.e. \( E(R(X)) = E(RES) \).

In order to estimate reserving margins we would estimate the distribution of the stochastic variable \( R + (RES - E(RES)) \).

**Example 2**

Again we consider the distribution outlined in example 1 and the reserving method \( R_f \) defined above. Since the claims are independent it follows that \( R(X) \) and \( (RES - E(RES)) \) are independent and we therefore only have to calculate the convolution of the distributions of \( R_f(X) \) and \( (RES - E(RES)) \), for example by simulation.

The distribution of \( R_f(X) \) is approximated by the Bootstrap distribution \( R_f(Y) \). The distribution of \( RES - E(RES) \) could have been estimated by simulation on the basis of the estimated parameters and the model assumptions. However, the distribution has been simulated on the basis of the original parameters. The results are outlined in figure 9 below.
It is seen, that the main contribution to the uncertainty of 'Total' rises from the estimation uncertainty. For example, the 95% fraction is only increased from 2622 to 2632 when the future randomness is included. It is often seen, as in this example, that focus should be on the randomness in the past rather than in the future when reserving margins are estimated.

**Application 3: Selection of reserving method**

We consider two different reserving methods $R_1$ and $R_2$. Assume that both methods are unbiased estimators i.e.

$$E(R_1(X)) = E(R_2(X)) = E(RES)$$

We would then prefer to use $R_2$ rather than $R_1$ if

$$\text{Var}(R_2(X)) < \text{Var}(R_1(X)).$$

The problem is that the distributions of $R_1(X)$ and $R_2(X)$ are unknown. However, using Bootstrap approximation, we can easily estimate the variances of $R_1(Y)$ and $R_2(Y)$ and base the selection on these.

**Example 3**

We consider the distribution outlined in example 1 and two different reserving methods:

$R_1$: Deterministic Chain Ladder method based on the payments (as above)

$R_2$: A stochastic model with unknown but constant claim inflation.

Both methods are reasonably central/unbiased, $E(R_1(X)) = 2130$, $E(R_2(X)) = 2145$ and $E(RES) = 2143$. 

<table>
<thead>
<tr>
<th></th>
<th>$R_1(Y)$-Bootstrap</th>
<th>RES - E(RES) estimated</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>2253</td>
<td>0</td>
<td>2253</td>
</tr>
<tr>
<td>STD</td>
<td>214</td>
<td>77</td>
<td>230</td>
</tr>
<tr>
<td>Median</td>
<td>2247</td>
<td>-4</td>
<td>2248</td>
</tr>
<tr>
<td>75% fraction</td>
<td>2373</td>
<td>39</td>
<td>2380</td>
</tr>
<tr>
<td>95% fraction</td>
<td>2622</td>
<td>120</td>
<td>2632</td>
</tr>
<tr>
<td>99% fraction</td>
<td>2713</td>
<td>166</td>
<td>2758</td>
</tr>
</tbody>
</table>
It is seen from figure 10 that the stochastic model $R_2$ gives a more precise reserve estimate but furthermore, the standard deviation of $R_2$ is less than that of the Chain Ladder method $R_1$:

<table>
<thead>
<tr>
<th></th>
<th>$R_1(Y)$ - Bootstrap</th>
<th>$R_1(X)$ - Real dist.</th>
<th>$R_2(Y)$ - Bootstrap</th>
<th>$R_2(X)$ - Real dist.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>2253</td>
<td>2130</td>
<td>2134</td>
<td>2145</td>
</tr>
<tr>
<td>STD</td>
<td>214</td>
<td>179</td>
<td>177</td>
<td>153</td>
</tr>
<tr>
<td>Median</td>
<td>2247</td>
<td>2118</td>
<td>2137</td>
<td>2144</td>
</tr>
<tr>
<td>75% fraction</td>
<td>2373</td>
<td>2229</td>
<td>2257</td>
<td>2244</td>
</tr>
<tr>
<td>95% fraction</td>
<td>2622</td>
<td>2456</td>
<td>2414</td>
<td>2418</td>
</tr>
<tr>
<td>98% fraction</td>
<td>2713</td>
<td>2508</td>
<td>2505</td>
<td>2487</td>
</tr>
</tbody>
</table>

Conclusion

It is concluded that if the individual claim data are available the Bootstrap is an effective method to calculate approximations to the uncertainty distributions of claims reserves. Only distribution assumptions regarding the claim frequency are required. The application is not dependent on the reserving method and even for complicated reserving methods the uncertainty can be estimated. It can be used to select between different reserving methods to obtain the most robust method and to calculate reserving margins.

References
