Monte Carlo Simulation Approach to Assess Health Care Claims Risk

Zhirong Mao, Blue Cross of California, San Diego, CA

ABSTRACT

This paper utilizes statistical functions of the SAS® system to present a simulation approach to approximate the probability distribution of claims, which is determined jointly by several random variables, such as health care access, average length of stay and average cost per unit. The simulated distribution provides an effective tool to assess the claims risk exposure without deriving the claims probability distribution in analytical form. Use of this approach is extensible to risk management, insurance product pricing and provider network development.

INTRODUCTION

Assessing claims risk is a key factor in health care plan design, actuarial pricing analysis and provider network development. To project the amount of claims, usually, three crucial variables are monitored: access, average length of stay (ALOS), facility or provider rates. Access is usually measured by patients per 1,000 members; average length of stay is measured by the number of hospital days for inpatient care or by the number of visits for outpatient care; rates are unit prices of service, one night hospital stay or one office visit. The average cost of care is thus determined by the three variables. Since all the three factors are random variables following certain probability distributions, calculating the average amount of claims is not sufficient for completely assessing the claims risk. Another important measurement needed is the variations of claims. However, the complexity of a joint probability distribution function of these three variables makes it impractical to analytically derive a probabilistic model of claims. Using statistical functions in the SAS system, I propose a simulation approach to assess the risk of claims. The simulated distribution, generated with random variables from three probability distributions, gives an effective and graphic tool to assess the claims risk exposure.

A SIMPLE SIMULATION

In the following simplified simulation for inpatient care, I start with the simple expression of claims,

\[ \text{Claims} = \text{Access} \times \text{ALOS} \times \text{Rate}, \]

and simulate the distribution of CLAIMS with ACCESS ALOS and RATE from three probability distributions. The variable of access to care (ACCESS) follows a Poisson distribution with a mean of 10 hospital admissions per 1,000 members. Average length of stay (ALOS) is a lognormal variate with a mean of nine days hospital days per hospital admission and a standard deviation of 6 days. The rate variable (RATE) comes from a normal distribution with a mean of 400 dollars and a standard deviation of 40. In the following SAS program, I have generated one thousand claims records. To simplify the code, I start with zero as the seed and use the loop count number, \( I \), as dummy seeds for random number generators. More discussions of seeds for random number can be found in Clark and Woodward (1992) and SAS Language Guide.

```
%Macro simu(a_lamda, I*Access average *11*1,000 member *11*lambda in Poisson *JIn_m, I*mean of LOS *JIn_sd, I*SD of LOS *11*in lognormal *Jrate_m, I*mean of Rates *Jrate_sd, I*SD of Rates *11*in normal dist. *Jcount); /*No. of iterations*/
data one;
do i=1 to &count;
   ACCESS=rannor(0,10);
   /* find mean (mu) and variance (sigma) */
   /* of normal dist. with specified mean*/
   /* and SD of lognormal dist. */
   sigma=log((&In_sd*&In_sd/&In_m/&In_m)+1);
   mu=log(&In_m)-sigma/2;
   ALOSI=exp(mu+sqrt(sigma)*rannor(i));
   RATE=rate_m+rate_sd*rannor(i);
   output;
end;
data two; set one;
claims=A1000*ALOS*R12;
proc print data=two (obs=100);
%Mend simu;
%simu(10,9, 7, 400, 40,1000);
```

Table 2.1
Simulation Output
First 10 observations with 1,000 iterations

<table>
<thead>
<tr>
<th></th>
<th>ACCESS</th>
<th>ALOS</th>
<th>RATE</th>
<th>CLAIMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>5</td>
<td>416.35</td>
<td>1.01</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>3</td>
<td>375.44</td>
<td>1.22</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>10</td>
<td>396.54</td>
<td>3.06</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>5</td>
<td>439.91</td>
<td>0.89</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>7</td>
<td>397.15</td>
<td>2.21</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
<td>5</td>
<td>377.36</td>
<td>1.89</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>17</td>
<td>425.66</td>
<td>6.21</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>13</td>
<td>357.63</td>
<td>1.88</td>
</tr>
<tr>
<td>9</td>
<td>13</td>
<td>15</td>
<td>371.29</td>
<td>6.21</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>12</td>
<td>444.52</td>
<td>3.19</td>
</tr>
</tbody>
</table>
Chart 1 Distribution of Access to Care
Poisson Distribution with $\lambda = 10$
Iterations = 1,000

Chart 2 Distribution of Average Length of Stay
Lognormal Distribution: Mean=9 days; SD=7 days
Iterations = 1,000

Chart 3 Distribution of Hospital Rates
Normal Distribution: Mean=$400 per day SD=40$
Iterations = 1,000

Chart 4 Distribution of Claims (PMPM)
Simulated Distribution
Iterations = 1,000

DISCUSSION

I selected Poisson, Lognormal and Normal distributions respectively for ACCESS, ALOS and RATE to reflect the following empirical findings: 1) the likelihood of access is proportional to the length of time interval; 2) Experience-based data shows that average length of stay tend to be non-center distributed and is close to the properties of lognormal distribution. 3) Hospital rates are structured with certain degree of variation and symmetry. The distributions used here are merely for illustration purposes. The SAS software has a broad range of statistical functions and procedures to solve more complex stochastic problems.

Chart 1 shows a Poisson distribution of ACCESS with an average of 10 admissions per 1,000 members. The shape of the Poisson probability function depends on the mean, $\lambda$, due to the fact that the distribution has both its mean and variance equal to $\lambda$. The variance of the distribution increases with its mean. This implies that higher access to care on average must be associated with greater dispersion of access to care, i.e., higher likelihood of extremely lower or higher access. Chart 5 has Poisson distributions with means from 0.4 to 4. Their shapes change significantly from one to another.

The average length of stay (ALOS) as illustrated on Chart 2 is characterized by a lognormal distribution with a mean of 9 and a standard deviation of 7. Notice the parameters, $\mu$ and $\sigma$, are calculated in the SAS program to generate a lognormal distribution with specified mean and variance. Relatively high frequencies are associated with the number of days less than 8. However, some long length of stay observations appear with low frequency. This captures quite well the empirical finding that most of patients have short and effective treatment at inpatient acute facilities while a small portion of patients with more serious illness stay much longer in hospitals. Actually, the longer-stay-less-frequent observations of patients with serious health problems form the “long-tail” on the chart and become a major component in the claims risk structure.
Chart 5: Poisson Distribution with Different \( \lambda \) Values

\[ \Pr(X) = \frac{e^{-\lambda} \lambda^x}{x!} \]

Iterations = 1,000

<table>
<thead>
<tr>
<th>( \lambda = 0.5 )</th>
<th>( \lambda = 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Pr(X)</td>
<td>Pr(X)</td>
</tr>
<tr>
<td>0.0  0.1  0.2  0.3</td>
<td>0.0  0.2  0.4  0.6</td>
</tr>
<tr>
<td>0.4  0.5  1.0  2.0</td>
<td>0.0  0.2  0.4  0.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \lambda = 2 )</th>
<th>( \lambda = 4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Pr(X)</td>
<td>Pr(X)</td>
</tr>
<tr>
<td>0.0  0.1  0.2  0.3</td>
<td>0.0  0.3  0.6  1.0</td>
</tr>
<tr>
<td>0.2  0.4  0.8  1.6</td>
<td>0.0  0.3  0.6  1.0</td>
</tr>
</tbody>
</table>

Chart 4 has rates from a normal distribution with a mean of $400 per day and a standard deviation of 40. This variable is usually more subject to human intervention, such as provider network development and contracted rates.

The most interesting and useful tool is the simulated distribution of claims on Chart 4. It is apparent that claims per member per month (PMPM) are very likely under $8 with a chance of 95% and claims PMPM can be under $3 with a 50-50 chance. Claims (PMPM) are now associated with probabilities. For instance, claims (PMPM) less than $2 have a probability of 42.5%; it does not make significant difference in terms of probability (close to 100%) after the claims (PMPM) exceed $9. This chart gives us a more manageable tool to assess the claims risk. The projection of claims is thus no longer just an idea of average. The dispersion of claims become visible and measurable. If we treat the claims dispersion around the mean as a kind of risk, it is understandable from Chart 4 that a riskier health plan would have a "longer-and-fatter-tail" probability distribution on Chart 4. This implies that "extreme" events, such as hits of high-cost claims, are more likely to happen.

APPLICATIONS

The method presented above can be applied, but not limited, to several areas, such as provider network development, case management, actuarial pricing and product design. Provider network development can adjust the parameters of RATES and simulate the impact on the financial outcome, claims. The analysis become bi-dimensional, embodying average contracted rates and the deviations of the rate structure. The simulation approach is capable of addressing some extreme scenarios, for instance, all seriously sick patients happen to be in the most expensive area and stay the longest in the hospitals. Case management follows certain practice guidelines. Adapting new practice guidelines such as reducing or extending length of stay would change the risk structure of claims. Actuarial pricing globally assess the product risk, administrative cost and other factors. Apparently, claims risk is a major component of the financial risk. To have a competitive and "safe" pricing, actuarial analysis needs to have both experience-based and estimation-based risk analysis. The simulation approach here would provide a claims risk gauge, which would facilitate the estimation-based analysis. Product design can adjust the parameters in the simulation model to estimate the claims risk involved in the new design. The approach presented in this paper provides only a simple framework for analyzing claims risk, which is capable to include other variables. More sophisticated simulation models need to incorporate benefit calculation algorithms (Mao and Xu, 1997) and other modeling techniques.

CONCLUSION

Simulation using the SAS System appears to be an effective and easy approach to assess health care claims risk. This paper only scratches the surface of the subject of applying simulation method in health care industry. To have meaningful simulation results, it is necessary to have a better understanding of the random variables involved in the simulation model and their counterparts in real practice.

ACKNOWLEDGMENTS

I benefit considerably from my discussions with colleagues at Blue Cross of California. Special thanks go to Carrie Striebeck and Shari Johnston-O’Neill for their insightful suggestions on improvements to the manuscript. This paper presenting a simulation approach in the SAS system does not reflect any health care policy or practice of Blue Cross of California. Any errors in this paper are solely my responsibility.

REFERENCES


Author

Zhirong Mao

Blue Cross of California
9655 Granite Ridge Drive, 6th Fl.
San Diego, CA 92123
Tel: (619) 571-8137