PRICING AN INSURANCE POLICY USING MONTE CARLO SIMULATION AND INTEGRATION

AJKANA HIMA 31033
INFOTRONICS SYSTEM ENGINEERING
RHEIN WAAL UNIVERSITYOF APPLIED SCIENCES

ABSTRACT

This research looks into the use of Monte Carlo simulation as a technique of pricing insurance premiums. The simulation-based technique provides a thorough and probabilistic estimation of policy prices by taking into consideration numerous parameters such as claim amounts, claim frequencies, and deductibles. The study generates random samples for claim amounts and frequencies using a dataset comprising various attributes such as age, gender, BMI, number of children, smoking status, region, and charges. The total claims for each insurance are determined by applying the deductible over several iterations of the simulation, resulting in a distribution of policy costs. To examine the uncertainty and precision of the estimates, statistical measures such as the average policy price, standard deviation, and confidence intervals are used. The data show that claim amounts, frequency of claims, and deductibles have a substantial impact on anticipated policy prices. In the context of insurance pricing, understanding the implications of anticipated policy costs is critical. The average policy price, coupled with the uncertainty range, gives useful information about the prospective expenses connected with the insurance policy. It enables insurance companies to make well-informed judgments about pricing and risk management. In conclusion, the use of Monte Carlo simulation in insurance policy pricing provides a robust and adaptive approach. The study emphasizes the significance of taking into account different aspects and applying uncertainty analysis into pricing models. Further research can concentrate on the development of advanced models, the use of larger datasets, and validation procedures to improve the accuracy and reliability of policy pricing estimates. This study adds to the body of knowledge in the subject of insurance pricing by proving the utility and potential of Monte Carlo simulation in creating realistic and informative policy prices.

TABLE OF CONTENTS

Introduction

What is Monte Carlo Simulation and Integration

Monte Carlo simulation is a computing approach that simulates random variables and their interactions to describe and evaluate complicated systems or processes. It is especially beneficial when obtaining analytical or deterministic answers is difficult or impossible. The method entails creating a huge number of random samples or situations based on input or probability distributions. These samples are then utilized to predict the behavior or result of the system or process being studied.

Monte Carlo integration is a numerical approach for estimating definite integrals using random sampling in the context of integration. It is based on the premise that the integral of a function can be approximated by multiplying the function's average value by the size of the integration domain. The average value can be approximated and multiplied by the domain size to give an approximation of the integral by generating random locations within the integration domain and evaluating the function at those points.

Monte Carlo simulation and integration are frequently utilized in industries such as finance, engineering, physics, and statistics. They provide a versatile and powerful method for studying and addressing problems involving uncertainty, unpredictability, or complex systems. These methods provide probabilistic estimates, enable sensitivity analysis, and aid in decision-making processes where analytical answers are not viable or practicable by simulating a large number of situations or sampling from a distribution.

Adjustment of this topic to my project:

The goal of this research is to estimate the cost of an insurance policy using Monte Carlo simulation and integration. We can simulate a large number of policy scenarios and evaluate the expected policy price by utilizing real-world data on claim amounts and claim frequencies. The Monte Carlo method incorporates uncertainty and unpredictability into the pricing process, resulting in more precise estimation.

The dataset utilized in this study comprises historical information on insurance claims. It has two important variables: claim amounts and claim frequencies.

Claim Amounts: The financial impact of individual insurance claims is represented by this variable. It keeps track of the monetary amount of each claim, indicating the cost incurred by the insurance company for that particular claim. Claim amounts might vary greatly based on factors such as the covered asset's nature or the type of coverage.

Claim Frequencies: This statistic represents the number of claims filed within a specific policy period. It gives information on how frequently policyholders submit claims for reimbursement or compensation. Claim frequencies might vary depending on factors such as the insured asset's risk profile, insurance terms, or external factors such as environmental conditions.

The dataset is derived from real-life insurance data and represents policyholder claims over a certain time period. It gives the information required to comprehend the historical patterns and characteristics of insurance claims, allowing us to effectively estimate the policy price.

Using Monte Carlo simulation techniques, we may simulate alternative scenarios by examining the claim amounts and claim frequencies in the dataset. This enables us to model the uncertainty and randomness associated with insurance claims and get a policy pricing estimate that takes these aspects into account.

Methodology

- Data Preparation: The insurance dataset, which includes information about policyholders such as age, gender, BMI, number of children, smoking status, region, and charges, is loaded.
- Claim Frequency and Amount Sampling: The code extracts the frequency and amount of claims from the dataset. These figures indicate the number of claims filed and the amounts paid for each policyholder.

- Deductible Amount: The deductible amount is specified. It reflects the limit below which the insurance company will not pay claims.
- Monte Carlo Simulation: To estimate the policy price, the program runs a Monte Carlo simulation. It loops through each policy, performing the following procedures for each simulation.

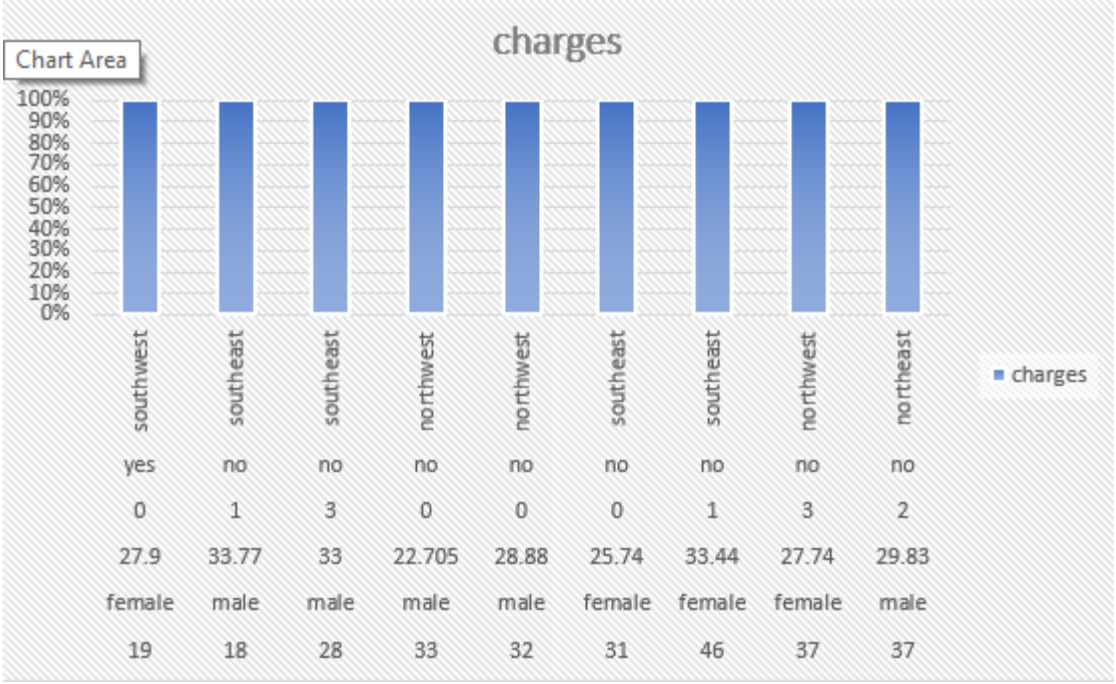
a) *Claim Calculation*: For each insurance, the claim frequency and claim amount are drawn at random from the dataset.

b) *Calculation of Total Claims:* The total claims are computed by multiplying the claim frequency by the claim amount.

c) *Calculation of Policy Price*: The policy price is computed as the maximum of total claims minus the deductible. If the outcome is negative, the value is set to zero.

d) *Policy Price Storage:* Each simulation's policy price is saved in the policy_prices list.

- Statistics Calculation: After all simulations are completed, the code calculates statistics to summarize the policy prices obtained. The average policy price, standard deviation, and 95% confidence interval are calculated using NumPy functions.
- Results Presentation: The code prints the estimated policy price, standard deviation, and 95% confidence interval to provide insights into the policy pricing estimation.



[The mathematical explanation]

The key mathematical concepts involved in the code are:

Monte Carlo Simulation: The code uses a Monte Carlo simulation technique to estimate the policy price. In each simulation, claim amounts and claim frequencies are randomly sampled from the provided dataset, and the policy price is calculated based on these sampled values.

Claim Calculation: For each simulation, the code calculates the total claims by multiplying the claim amount and claim frequency. This represents the total amount that the insurance company would need to pay out for that particular policy.

Deductible: The code applies a deductible to the total claims. The deductible is subtracted from the total claims, and if the result is negative, it is set to zero. This ensures that the policy price is not negative.

Average Policy Price: The code calculates the average policy price by taking the mean of the policy prices obtained from all simulations. This provides an estimate of the expected policy price based on the Monte Carlo simulation.

Standard Deviation: The code calculates the standard deviation of the policy prices obtained from all simulations. The standard deviation measures the variability or spread of the policy prices, indicating the degree of uncertainty in the estimates.

Confidence Interval: The code calculates the 95% confidence interval for the policy prices using the np.percentile() function. The confidence interval provides a range within which the true average policy price is expected to fall with a 95% level of confidence.

By using these mathematical calculations and the Monte Carlo simulation approach, the code provides estimates for the policy price, measures of uncertainty (standard deviation), and a confidence interval for the insurance policy pricing problem.

Data Description

The insurance dataset utilized in this Python program contains information about insurance policyholders and the healthcare bills they incur. The variables in the dataset are as follows:

1. Age: The age of the policyholder (numeric).
2. Gender: The gender of the policyholder (categorical: "female" or "male").
3. BMI: The body mass index of the policyholder (numeric).
4. NumChildren: The number of children the policyholder has (numeric).
5. Smoker: Indicates whether the policyholder is a smoker or not (categorical: "yes" or "no").
6. Region: The region where the policyholder resides (categorical: "southwest", "southeast", "northwest", or "northeast").
7. Charges: The healthcare charges incurred by the policyholder (numeric).

The dataset contains data on individual policyholders, their characteristics, and the healthcare expenditures they incurred.

age	sex	bmi	children	smoker	region	charges
19	female	27.9	0	yes	southwest	16884.92
18	male	33.77	1	no	southeast	1725.552
28	male	33	3	no	southeast	4449.462
33	male	22.705	0	no	northwest	21984.47
32	male	28.88	0	no	northwest	3866.855
31	female	25.74	0	no	southeast	3756.622
46	female	33.44	1	no	southeast	8240.59
37	female	27.74	3	no	northwest	7281.506
37	male	29.83	2	no	northeast	6406.411

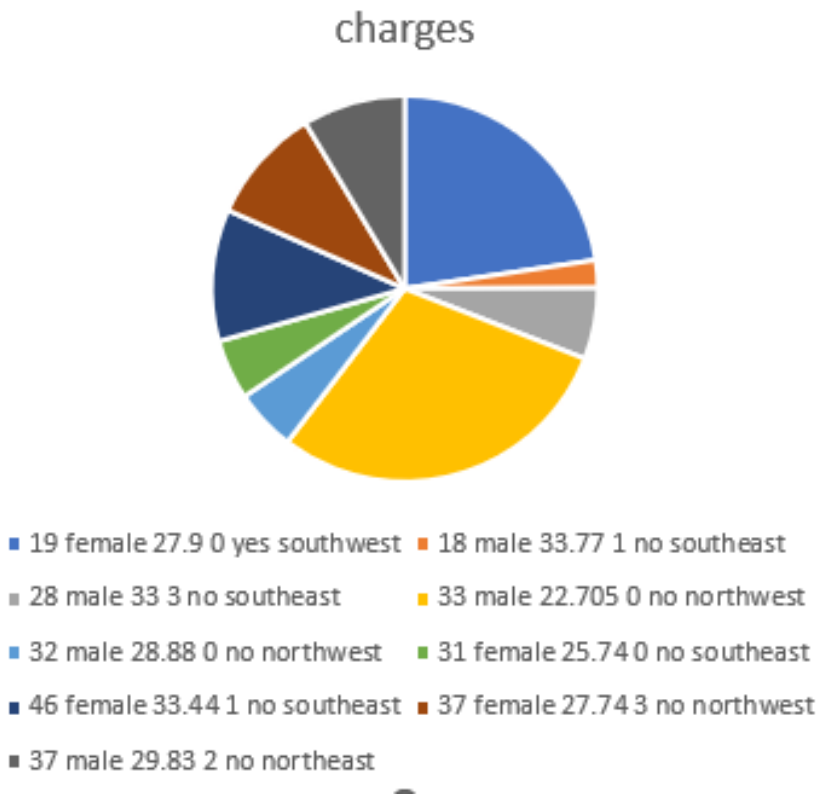
Sources and Limitations

In the provided context, the source of the data used in this project is not disclosed. Insurance companies, on the other hand, frequently collect such data from their policyholders for real-world uses.

Some constraints and assumptions that may apply to the dataset and the project are as follows:

The dataset may reflect a small sample of policyholders and healthcare charges that does not fully represent the total population.

Missing Data: The dataset may have missing data, which could affect the analysis and conclusions. This code assumes that the dataset is complete and that there are no missing values.



Generalizability: The findings from this dataset may not be applicable to all insurance policyholders or insurance marketplaces. The dataset and analysis are based on specific charges and characteristics observed in the submitted dataset.

Assumptions: The code assumes that the provided dataset reliably represents policyholder claim rates, claim amounts, and other variables. It also presupposes that the Monte Carlo simulation approach appropriately captures the link between the variables and policy pricing.

Conclusion

Monte Carlo simulation provides a powerful and adaptable method for estimating insurance prices by taking into account numerous claim scenarios and adding uncertainty. It enables insurance companies to account for the stochastic character of claim frequencies and amounts, resulting in a more accurate risk and pricing evaluation.

The Monte Carlo simulation gives a range of possible policy prices by simulating different scenarios, allowing insurers to estimate not only the average policy price but also measures of uncertainty such as confidence intervals and standard deviations. This information is critical for making decisions, managing risks, and determining suitable premiums.

One area for future research and development is to look at more sophisticated models that can account for additional factors influencing policy prices. This includes factors such as policyholder characteristics, market dynamics, and external variables such as economic data. By taking these elements into account, the models can provide a more thorough and accurate assessment of policy prices while accounting for diverse affecting factors.

Another area to focus on is data enrichment. This involves incorporating more comprehensive and detailed datasets into the simulation. This can include increasing the sample size, extending the time period of the data, and including more diverse policyholder profiles. By enriching the data, the simulations can provide more robust and reliable estimates of policy prices, considering a wider range of scenarios and variations.

Conducting sensitivity analysis is important to assess the impact of different assumptions and variables on policy prices. By varying the inputs and assessing how changes in underlying factors affect the results, insurers can gain insights into the sensitivity and stability of the pricing estimates. This can help in identifying critical variables and understanding their influence on the policy prices.

Validating the simulation results against historical data or benchmarking against industry standards is essential to ensure the accuracy and reliability of the pricing estimates. By comparing the simulation outputs with real-world data or industry

benchmarks, insurers can validate the effectiveness of the simulation model and make necessary adjustments or calibrations to improve its accuracy.

In summary, further research and improvement in the methodology of insurance policy pricing can include exploring enhanced models, incorporating comprehensive datasets, conducting sensitivity analysis, and validating the simulation results. These efforts can lead to more accurate and reliable pricing estimates, better understanding of the factors influencing policy prices, and improved decision-making in the insurance industry.