Econometrics of Insurance

Mai-Anh Dang | Student ID: 21608631

April 26, 2018

Abstract

This project would be separated into 02 parts. In the first part, the actuarial premium for policyholders would be estimated, conditioned on the characteristics of drivers, cars, and relevant policy in contracts. The variables of interests are the frequency of claims and the severity (paid amount) of claim. We would estimate the former with Poisson and Negative Binomial models, and the latter with Gamma and Log-likelihood specifications. Models would be compared and discussed to choose the more appropriate. Afterwards, the priori and posteriori fair premium would be computed. This project is based on the data sets of *Actuarial Pricing Game, 2017.* [1]

Part I

Priori Pricing

1 Merging Data

There are two data sets used in this project:

- `PG_2017_CLAIMS_YEAR0`: This data includes the claims of clients in Year 0. The interest variable is `claim_nb`, originally taking value of 1 for each claim. The `claim_amount` is the individual claim amounts for each case, range from -2,000 to +300,000. There are negative values when the driver's liability is not engaged and there is a legal recourse. Each claimed case is identified by `id_client` and `id_vehicle` (under the fact that an individual could own several cars then each observations are treated independently).
- `PG_2017_YEAR1`: In this data set, corresponding with `id_client` and `id_vehicle`, we have other variables about the relevant policy, characteristics of drivers and vehicles.

In `PG_2017_CALIMS_YEAR0`, each client with each car might claim for several times, which leads to the duplicated key `id_client * id_vehicle`. This burden the merging two data sets. Thus, first of all, we sum up all times and claim amounts for each `id_client * id_vehicle`. Then, we merge two data sets (by left join) [2]. In the merged data set, we have 100,000 observations with 33 variables. Descriptive Statistics of important variables is in **Table 2**.

The histogram of `claim_nb` and `log(claim_amount)` is displayed in **Figure 1**. We can see that the most clients have less than 1 time of claim. In fact, the number of clients by numbers of claims is as below:

claim_nb	0	1	2	3	4	5	6
Obs.	87,346	11,238	1,264	134	16	1	1

Table 1: Number of observations by number of claims

[1] http://freakonometrics.free.fr/PG3/toolbox.zip

[2] The merged data set is named `merged_final_insurance.csv` included in the submission of this report

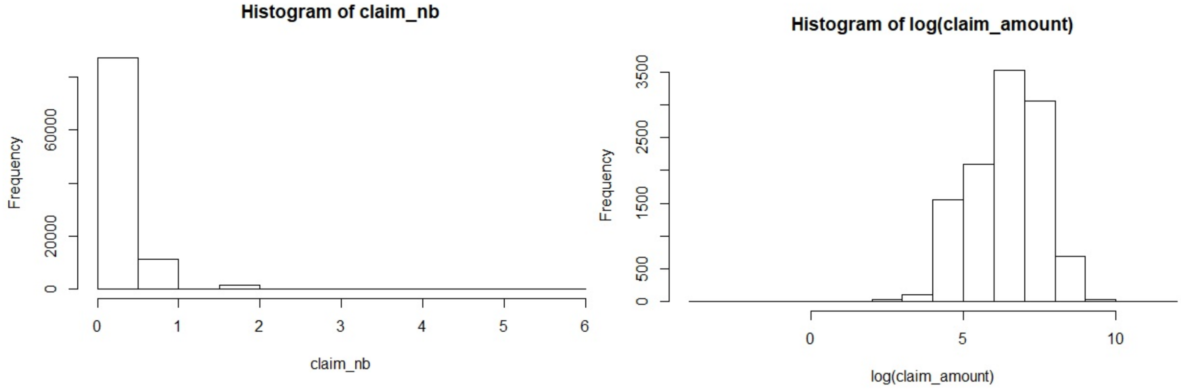

Figure 1: Distribution of claim numbers and amounts

Table 2: Descriptive Statistics

	Mean	Std.	Median	Min	Max
claim_nb	0.140	0.400	0.000	0.000	6.000
claim_amount	117.25	926.08	0.000	-1863.92	141828.62
pol_bonus	0.530	0.090	0.500	0.500	1.950
pol_coverage	1.700	1.040	1.000	1.000	4.000
pol_duration	12.080	8.550	10.000	2.000	42.000
pol_sit_duration	3.730	2.360	3.000	2.000	26.000
drv_age1	55.680	14.870	55.000	20.000	104.00
drv_age2	15.91	24.44	0.000	0.000	100.00
drv_sex1	1.600	0.490	2.000	1.000	2.000
drv_sex2	1.460	0.710	1.000	1.000	3.000
drv_age_lic1	33.490	13.470	34.000	2.000	112.00
drv_age2_lic2	9.460	17.360	0.000	0.000	112.00
vh_age	10.550	7.030	9.000	2.000	67.000
vh_cyl	1645.88	461.93	1587.00	0.000	6997.00
vh_din	91.390	34.310	87.000	13.000	555.00
vh_sale_begin	12.650	7.790	11.000	2.000	75.000
vh_sale_end	9.620	6.700	8.000	1.000	56.000
vh_speed	170.68	23.370	170.00	25.000	310.00
vh_value	18058.69	8663.27	16229.5	0.000	155498.00
vh_weight	1128.21	360.64	1130.00	0.000	7901.00

2 Estimate the Frequency Claim

In this project, we would estimate the frequency claim by generalized regression models for count data, with **Poisson Distribution** and **Negative Binomial Distribution**.

2.1 Poisson Distribution

In the *Poisson regression model*, the dependent variable is `claim_nb`, as random variable $\mathbf{Y}$, which shows the number of claims per one insurance policy in a vehicle insurance portfolio. It would be model through *Poisson*(λ) distribution. The Poisson distribution indicates number of events (insurance claims) in a specified time interval.

The probability mass function is:

$$Y \sim Poisson(\lambda), \text{ and } f(y;\lambda) = \frac{\lambda^y}{y!}e^{-y} \text{ with } y \in N, \lambda > 0 \tag{1}$$

The expectation and variance: $E[Y] = Var[Y] = \lambda$
For the sample variable, when $Y_i \sim Poisson(\lambda_i)$, then $\mu_i = E[Y_i] = \lambda_i$. We have the link function (to link the random variable and systematic component) as below:

$$\eta_i = ln(\lambda_i) \text{ , when } \eta_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} \tag{2}$$

$$\lambda_i = d_i exp\left(\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\right) \tag{3}$$

From (2) and (3), the Poisson regression model for the estimation of claim frequency is:

$$\ln \lambda_i = \ln d_i + \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} \tag{4}$$

where $\ln d_i$ is the natural logarithm of risk exposure, or *"offset variable"*. d_i, representing the risk exposure of insurance policy i, namely the time interval, usually expressed in years, from the initial moment wen the policy was issued until the moment the sample is observed.
In fact, the model does not consider insurance policy equivalent in time interval (for example, some policies have longer duration). In other words, the Poisson model handle rates as a count per unit time. In this case, the exposure variables is necessarily involved to change the dependent variable from a rate to a count. By adjusting the offset variable, we adjust the scale of opportunities of an events. Intuitively, we can think that when each year is simply an opportunity for an claim to happen. Then, a duration of 20 years would be twice as likely to have a clam than a duration of 10 years. Under the assumption that the likelihood of events is not changing over time.

The Poisson models would be estimated with the offset variables as below:

- **Column (1) - without offset variable**
- **Column (2)** - `pol_duration`: indicates how old the policy is, in years.
- **Column (3)** - `pol_sit_duration`: indicates how old the current policy characteristics are, which can be different from `pol_duration` as the same insurance policy might evolved in the past (changing coverage, vehicle, drivers, etc.)

The results of Poisson models are reported in **Table 3**. In which, according to AIC and and Log-Likelihood, the model with `pol_duration` is the best one. Models with offset variable is relatively better than the baseline model with no offset variable.
One can see that indeed there is some differences in the estimates using different offset variable. For example, for `pol_payd` indicates the subscription of the mileage-based policy ("Pay as you go"), while model (1) estimates the negative sign. The model (2) with `pol_sit_duration` as offset variable indicates the positive significant value. It means that the subscription to mileage-based policy, in fact, increase the frequency of claims. It is quite reasonable when we keep in minds that the different between `pol_duration` and `pol_sit_duration` is that the latter takes into account the changes of each insurance policy by the time, and reflects how old the current policy characteristics are.
While in Model (1), the age of driver would increase the frequency of claim. The models with offset value shows the negative sign, or the older driver would have less accidents. This difference could be explained that for the older drivers, they would likely have a longer duration of policy then the higher risk exposure. If we do not take into account this exposure, the estimated result might be misleading.

Besides that, for other features, all models present quite similar in value and sign of the coefficients. The interpretation of results are intuitive and reasonable. For example, the more experienced driver 1, with higher `drv_age_lic` would have lower frequency of claims. Meanwhile, the similar effect is not significant in driver 2. The older vehicle would have lower frequency of claims. Vehicles with higher speed would have higher frequency of claims. Male drivers have significantly higher.

Table 3: Poisson Distribution Count Models

	Dependent variable:		
	claim_nb		
	(1)	(2)	(3)
pol_paydYes	−0.155***	−0.129***	0.442***
	(0.049)	(0.049)	(0.049)
pol_usageProfessional	−0.279	−0.252	−0.347*
	(0.195)	(0.195)	(0.195)
pol_usageRetired	−0.482**	−0.220	−0.406**
	(0.194)	(0.194)	(0.194)
pol_usageWorkPrivate	−0.459**	−0.438**	−0.486**
	(0.193)	(0.193)	(0.193)
drv_drv2Yes	0.155***	0.068	0.326***
	(0.052)	(0.053)	(0.053)
drv_age1	0.007***	−0.007***	−0.002
	(0.002)	(0.002)	(0.002)
drv_age2	−0.004***	−0.001	−0.006***
	(0.001)	(0.001)	(0.001)
drv_sex1M	−0.076***	−0.060***	−0.076***
	(0.020)	(0.020)	(0.020)
drv_sex2F	0.085***	0.072**	0.083**
	(0.032)	(0.032)	(0.032)
drv_age_lic1	−0.008***	−0.023***	−0.008***
	(0.002)	(0.002)	(0.002)
drv_age_lic2	0.001	−0.0002	0.0004
	(0.001)	(0.001)	(0.001)
vh_age	−0.040***	−0.035***	−0.066***
	(0.006)	(0.006)	(0.006)
vh_cyl	0.0002***	0.0002***	0.0002***
	(0.00004)	(0.00004)	(0.00004)
vh_din	−0.001	0.001**	0.0003
	(0.001)	(0.001)	(0.001)
vh_fuelGasoline	−0.178***	−0.198***	−0.227***
	(0.024)	(0.023)	(0.024)
vh_fuelHybrid	0.068	0.177	0.009
	(0.243)	(0.243)	(0.243)
vh_sale_begin	−0.018***	−0.018***	−0.026***
	(0.005)	(0.005)	(0.005)
vh_sale_end	0.001	−0.003	0.001
	(0.005)	(0.005)	(0.005)
vh_speed	0.003***	0.002***	0.003***
	(0.001)	(0.001)	(0.001)
vh_typeTourism	0.010	−0.103**	−0.030
	(0.041)	(0.041)	(0.041)
vh_weight	0.00002	−0.00003	0.00001
	(0.00003)	(0.00003)	(0.00003)
Constant	−1.774***	−2.959***	−2.124***
	(0.231)	(0.231)	(0.232)
Observations	100,000	100,000	100,000
Log Likelihood	−41,964.610	−45,727.690	−43,215.210
Akaike Inf. Crit.	83,973.220	91,499.390	86,474.430
Note:			*p<0.1; **p<0.05; ***p<0.01

2.2 Negative Binomial Distribution

The limitation of Poisson model is its assumption that mean and variance is equal: $E[Y] = Var[Y]$. In fact, it is usually observed that the conditional variance is larger than the conditional mean. It is particularly likely in the situation of unobserved heterogeneity. We would diagnostic that in the next section by the *Test of Overdispersion*.

In the case of unobserved heterogeneity, the alternative model is **negative binomial**. Different from the Poisson model which only relies on its mean $\mu_i = \lambda_i$, the negbi. replies on two parameters: its mean is μ, but the conditional variance is $\mu(1 + \alpha\mu)$.

The results of Negative Binomial models are presented in **Table 6**. One can see that the estimated coefficients are not much different comparing the results of Poisson models.

Include the duration for the model without the exposure

2.3 Diagnostics among Poisson and Neg.Bin.

Overdispersion Test This test is conducted by function `dispersiontest()` in `R`, package `AER`. It assess the null hypothesis that the assumption of Poisson holds:

$$H_0 : E[Y] = Var[Y] = \mu$$

$$H_1 : Var[Y] = \mu + \alpha \times \texttt{trafo}(\mu)$$

Overdispersion corresponds to: $\alpha > 0$. The coefficient of α could be estimated by an auxiliary OLS regression and test with t- or z-statistic, asymptotically standard normal under the null. The common specifications of the transformation function `trafo` are: $\texttt{trafo}(\mu) = \mu^2$ and $\texttt{trafo}(\mu) = \mu$. The former is negative binomial with quadratic variance and latter is negative binomial with linear variance function: $Var[Y] = (1 + \alpha)\mu = dispersion * \mu$. We test in both specifications, the results are in **Table 4**. In all cases, we always reject the null hypothesis (true α is likely greater than 0), then the Poisson model might not be appropriate due to the heterogeneity. The Negative Binomial model is preferred.

Table 4: Overdispersion Test

	`trafo`	z	p-val	Sample est.: alpha
Model 1	1	12.976	$< 2.2e-16$	0.0871
Model 2	1	16.547	$< 2.2e-16$	0.2570
Model 3	1	12.045	$< 2.2e-16$	0.1262
Model 1	2	12.045	$< 2.2e-16$	0.5267
Model 2	2	8.1652	$< 2.2e-16$	0.6929
Model 3	2	11.438	$< 2.2e-16$	0.6047

LR Test We would also conduct the LR test. The idea of the test is that NB and Poisson are nested that Poisson is a special case of NB, with the restriction that: $\alpha = 0$. A LR test would compare the two models to test if the $\alpha = 0$, or $\alpha > 0$ (on-sided test, it cannot be negative), then the Poisson is rejected against NB.

Table 5: LR Test

	$LR = -2(LL(Poisson) - LL(NB))$	p-val
Model 1	149.792	$1.924e-34$
Model 2	1,087.74	$1.521e-238$
Model 3	492.655	$3.768e-109$

The results of LR test are presented in **Table 5**. In all three set-ups of offset values, we reject the null hypothesis. This model would be used in the next steps of this project.

Table 6: Negative Binomial Count Models

	Dependent variable:		
	claim_nb		
	(1)	(2)	(3)
pol_paydYes	−0.153***	−0.131*	0.436***
	(0.051)	(0.069)	(0.056)
pol_usageProfessional	−0.274	−0.299	−0.358
	(0.217)	(0.293)	(0.241)
pol_usageRetired	−0.478**	−0.282	−0.424*
	(0.216)	(0.292)	(0.240)
pol_usageWorkPrivate	−0.454**	−0.491*	−0.502**
	(0.215)	(0.290)	(0.239)
drv_drv2Yes	0.155***	0.069	0.330***
	(0.055)	(0.075)	(0.062)
drv_age1	0.007***	−0.006***	−0.002
	(0.002)	(0.002)	(0.002)
drv_age2	−0.004***	−0.001	−0.007***
	(0.001)	(0.001)	(0.001)
drv_sex1M	−0.077***	−0.064**	−0.078***
	(0.021)	(0.029)	(0.024)
drv_sex2F	0.086**	0.077*	0.084**
	(0.034)	(0.046)	(0.038)
drv_age_lic1	−0.008***	−0.023***	−0.008***
	(0.002)	(0.002)	(0.002)
drv_age_lic2	0.001	−0.0004	0.0003
	(0.001)	(0.001)	(0.001)
vh_age	−0.041***	−0.037***	−0.064***
	(0.006)	(0.008)	(0.007)
vh_cyl	0.0002***	0.0002***	0.0002***
	(0.00004)	(0.0001)	(0.0001)
vh_din	−0.001	0.002*	0.0002
	(0.001)	(0.001)	(0.001)
vh_fuelGasoline	−0.175***	−0.206***	−0.223***
	(0.025)	(0.033)	(0.028)
vh_fuelHybrid	0.055	0.178	−0.013
	(0.263)	(0.356)	(0.295)
vh_sale_begin	−0.018***	−0.018**	−0.027***
	(0.006)	(0.008)	(0.006)
vh_sale_end	0.001	−0.003	0.001
	(0.006)	(0.007)	(0.006)
vh_speed	0.003***	0.002**	0.003***
	(0.001)	(0.001)	(0.001)
vh_typeTourism	0.012	−0.102*	−0.022
	(0.043)	(0.058)	(0.048)
vh_weight	0.00002	−0.00002	0.00002
	(0.00003)	(0.00005)	(0.00004)
Constant	−1.790***	−2.856***	−2.142***
	(0.253)	(0.342)	(0.282)
Observations	100,000	100,000	100,000
Log Likelihood	−41,889.710	−45,183.820	−42,968.890
Akaike Inf. Crit.	83,823.430	90,411.640	85,981.770
Note:			*p<0.1; **p<0.05; ***p<0.01

3 Estimate the Severity Claim

While the frequency model is to model the number of claims, using the characteristics of policies, drivers and vehicles, the severity model is to model the amount of claims, given a claim is incurred. Then, we could used the estimated frequency and severity to compute the expected total costs:

$$\text{Expected total costs} = \text{expected number of claims} \times \text{expected amount of one claim}$$

The total claim amount over the period: $S = \sum_{k=1}^{Y} C_k$. If severity is independent to frequency then the pure premium for the class of risk X:

$$E[S|X] = E[Y|X] \times E[C|X]$$

We would conduct the `glm` to estimate with different assumed distribution of C (amount of claim): (i) Gamma; and (ii) Log-normal. The results are displayed in **Column (1) and (2), Table 7** respectively.

3.1 Gamma Distribution

Gamma distribution: $C \sim \Gamma(\alpha, \beta)$ Density function:

$$f(c|\alpha, \beta) = c^{\alpha-1}\frac{\beta^{\alpha}e^{-\beta c}}{\Gamma(\alpha)}, \text{ with } c \geq 0; \alpha, \beta > 0$$

The mean and variance: $E[C] = \alpha/\beta$; $Var[C] = \alpha/\beta^2$

3.2 Log-normal Distribution

Log-normal distribution: $C \sim LogN(\mu, \sigma^2)$ if $W = ln(C) \sim N(\mu, \sigma^2)$ Density function:

$$f(c|\mu, \sigma^2) = \frac{1}{c\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln c - \mu)^2}{2\sigma^2}\right]$$

The mean and variance: $E[C] = E[e^Y] = e^{\mu+\sigma^2/2}$; $Var[C] = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2}$

3.3 Discussion and Compare Models

In terms of *Log Likelihood* and *AIC*, the Gamma distribution model is preferred rather than the Log-normal distribution model. In particular, this model has lower AIC and higher log likelihood. The Analysis of Variance (ANOVA) is also conducted, and presented in **Figure 3, Appendix A**, in which the observed variation s partitioned into component attributable to different source of variations. Also, we compare the residual deviance of two models.

The sign and values of coefficients are quite similar between two models. Comparing the reference of Biannual payment, the payment by monthly significantly has higher amount of claim. More experienced drivers likely cause less severe accident with significantly lower severity of claim. The time of being introduced in the market also has significant impacts on the severity, which is reasonable as the model and the trendiness would determine the cost of repairing or compensating these vehicles. For older cars, the amount of claim is lower. These results are reasonable and intuitive.

4 Compute the Premium

From the estimated frequency (by *Negative Binomial*) and the estimated severity (by *Gamma* and *Log Normal*), we compute the pure premium: $E[S|X] = E[Y|X] \times E[C|X]$. The results of computed premium is in the file attached with this report under the name `actuarial_premium.csv`. The distribution of estimated premium by Gamma and Log-normal model is displayed in **Figure 2**. One can see that estimates from Gamma model is flatter with a long and thicker tails, when Log-norm estimates peak around zero value. In our case, the former has better log-likelihood, which means that its distribution is closer the reality of data.

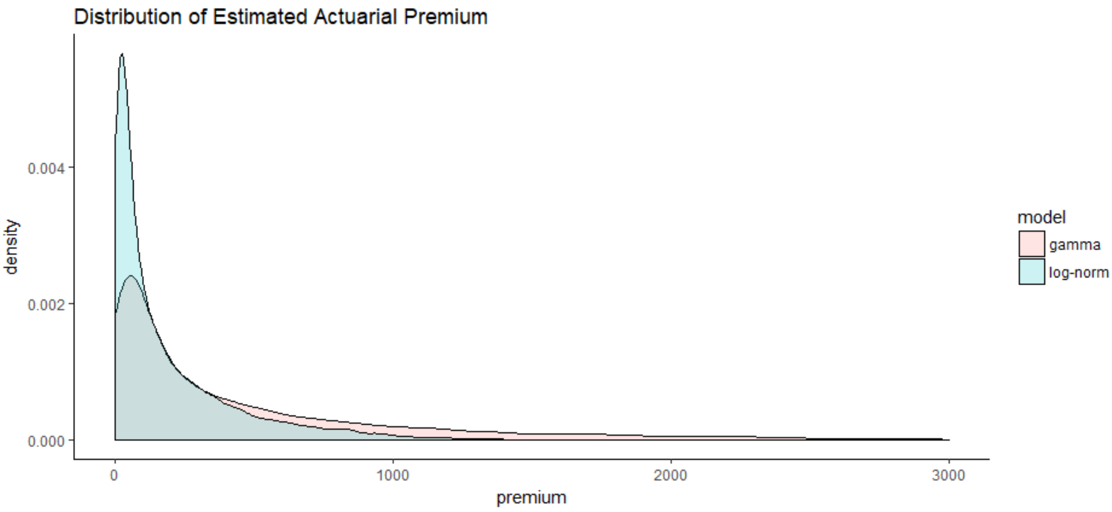

Figure 2: Estimated Premium by Gamma and Log-Norm Model

Table 7: Severity Claim Model: Gamma and Log-normal

	Dependent variable:	
	claim_amount	
	Gamma	*log-norm*
	(1)	(2)
pol_paydYes	−0.150	−0.099
	(0.215)	(0.173)
pol_pay_freqMonthly	0.326***	0.152**
	(0.095)	(0.067)
pol_pay_freqQuarterly	0.264	0.356**
	(0.220)	(0.148)
pol_pay_freqYearly	0.159*	0.008
	(0.092)	(0.067)
drv_age_lic1	−0.014***	−0.013***
	(0.003)	(0.002)
vh_age	−0.049**	−0.009
	(0.025)	(0.019)
vh_cyl	0.00001	−0.00004
	(0.0001)	(0.0001)
vh_din	0.004**	0.003**
	(0.002)	(0.001)
vh_sale_begin	0.049**	0.009
	(0.023)	(0.018)
vh_sale_end	−0.036	−0.013
	(0.023)	(0.017)
Constant	5.136***	4.496***
	(0.199)	(0.153)
Observations	11,102	11,102
Log Likelihood	−91,603.520	−103,195.000
Akaike Inf. Crit.	183,229.000	206,412.000
Note:	*p<0.1; **p<0.05; ***p<0.01	

Part II
Posteriori Pricing

Now we follow the assumptions that:

- 02 types of risk in the population, low and high-risk with 50% for each
- Number of claim N_{it} follows the Poisson Distribution with $\lambda_L = 0.05$ and $\lambda_H = 0.15$
- The realizations of severity of claims are i.i.d. and on average equal to 1
- Use *Bayesian Rule* to revise the *prior pricing*

5 Priori Fair Premium

The priori premium is:

$$Premium_{priori} = E[N_i]E[C_i]$$

In which, by the assumptions, $E[C_i] = 1$ and i.i.d, by the probabilities of high-risk and low-risk $Pr(L) = Pr(H) = 0.5$, we have:

$$\begin{aligned} Premium_{priori} &= E[N_{i1}] \\ &= E[N_{i1}|L]Pr(L) + E[N_{i1}|H]Pr(H) \\ &= \lambda_L \times 0.5 + \lambda_H \times 0.5 \\ &= 0.05 \times 0.5 + 0.15 \times 0.5 \\ &= 0.10 \end{aligned}$$

6 Posteriori Fair Premium

Table 8: Posteriori Fair Premium

k	$Pr(H\|N_{i1})$	$Pr(L\|N_{i1})$	$E[N_{i2}\|N_{i1}]$
0	0.475	0.525	0.098
1	0.731	0.269	0.123
2	0.891	0.109	0.139
3	0.961	0.039	0.146
4	0.987	0.013	0.149
5	0.995	0.005	0.150

The posterior fair premium:

$$E[N_{i2}|N_{i1} = k] = E[N_{i2}|N_{i1} = k, L] \times Pr(L|N_{i1} = k) + E[N_{i2}|N_{i1} = k, H] \times Pr(H|N_{i1} = k) \quad (5)$$

When the number of claims follow the Poisson Distribution and only depend on types, we have:

$$E[N_{i2}|N_{i1} = k, T] = E[N_{i2}|T] = \lambda_T, \text{in which} T \in \{L, H\}$$

The Equation (5) becomes:

$$E[N_{i2}|N_{i1} = k] = \lambda_L \times Pr(L|N_{i1} = k) + \lambda_H \times Pr(H|N_{i1} = k) \quad (6)$$

By Bayes Rule, we can compute $Pr(T|N_{i1})$ as below:

$$Pr(T|N_{i1} = k) = \frac{Pr(N_{i1} = k|T) \times Pr(T)}{Pr(N_{i1} = k|L) \times Pr(L) + Pr(N_{i1} = k|H) \times Pr(H)} \quad (7)$$

As $Pr(T) = Pr(L) = Pr(H) = 0.5$, we have:

$$Pr(T|N_{i1}) = \frac{Pr(N_{i1} = k|T)}{Pr(N_{i1} = k|L) + Pr(N_{i1} = k|H)} \tag{8}$$

$$= \frac{\exp(-\lambda_T)\frac{\lambda_T^k}{k!}}{\exp(-\lambda_L)\frac{\lambda_L^k}{k!} + \exp(-\lambda_H)\frac{\lambda_H^k}{k!}} \tag{9}$$

The results of this posteriori fair premium computation by $k = 1, \cdots, 5$ is presented in **Table 8**.

A Analysis of Deviance: Gamma and Log-Normal Severity Model

```
Analysis of Deviance Table

Model: Gamma, link: log

Response: claim_amount

Terms added sequentially (first to last)


              Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                          11101      21704
pol_payd       1    38.11     11100      21666 0.1078611
pol_pay_freq   3   260.78     11097      21405 0.0005089 ***
drv_age_lic1   1   337.69     11096      21068 1.697e-06 ***
vh_age         1   377.01     11095      20691 4.248e-07 ***
vh_cyl         1   189.70     11094      20501 0.0003338 ***
vh_din         1    71.86     11093      20429 0.0272367 *
vh_sale_begin  1    38.42     11092      20391 0.1064103
vh_sale_end    1    32.99     11091      20358 0.1346420
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> anova(lnorm.model, test = "Chisq")

Analysis of Deviance Table

Model: gaussian, link: log

Response: claim_amount

Terms added sequentially (first to last)


              Df  Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                           11101 7.7422e+10
pol_payd       1  14909520     11100 7.7407e+10 0.1428468
pol_pay_freq   3 126967893     11097 7.7280e+10 0.0003844 ***
drv_age_lic1   1 150158020     11096 7.7130e+10 3.318e-06 ***
vh_age         1  61581292     11095 7.7068e+10 0.0029023 **
vh_cyl         1  27940900     11094 7.7040e+10 0.0448683 *
vh_din         1  28060656     11093 7.7012e+10 0.0444120 *
vh_sale_begin  1    214676     11092 7.7012e+10 0.8604322
vh_sale_end    1   4207971     11091 7.7008e+10 0.4363123
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


COMPARE TWO MODEL

       Resid. Df Resid. Dev Df    Deviance Pr(>Chi)
Gamma      11091 2.0358e+04
Lnorm      11091 7.7008e+10  0 -7.7008e+10
```

Figure 3: ANOVA test: Gamma and Log-Normal