Ahn Vu Questions Not Yet Used

Anh Vu

August 4, 2019

These questions are not multiple choice types. Bring in later, possibly using **Datacamp**.

1 Chapter 8

1.1 Categorical variables and multiplicative tariff

1. Consider the data set "dataCar" in the R package "insuranceData". It contains the the number of claims (numclaims) and exposure (exposure) for individual policyholder. It also contains various rating factors including:

   - Vehicle body type (*veh_body*): BUS, CONVT, COUPE, HBACK, HDTOP, MCARA, MIBUS, PANVN, RDSTR, SEDAN, STNWG, TRUCK, UTE
   - Vehicle age (*veh_age*): 1 (youngest), 2, 3, 4
   - Gender (*gender*): F, M
   - Area (*area*): A, B, C, D, E, F
   - Policyholder's age (*agecat*): 1 (youngest), 2, 3, 4, 5, 6

   (a) Write down a Poisson regression model with log-link on the number of claims with respect to the above rating factors. For area, consider only 2 factors: area D and Others. Use base factors:vehicle body type SEDAN, vehicle age 1, gender F, area D, policyholder's age category 1.

   Answer:

$$\log \mu_i = \beta_0 + \sum_{t=1}^{12} \beta_t I(veh\_body = type_t) + \sum_{t=13}^{15} \beta_t I(veh\_age = t - 11)$$
$$+ \beta_{16} I(gender = M) + \beta_{17} I(area = Others) + \sum_{t=18}^{22} \beta_t I(agecat = t - 16) + \log m_i.$$

   For simplicity of expression, $type_t$, $t = 1, ..., 12$ represents the 12 types of vehicle type other than SEDAN.

   (b) Perform the Poisson regression in R.

   (c) What is the relativity in the tariff for SEDAN vehicle type?

   Answer: 1

   (d) What is the relativity in the tariff for area Other?

   Answer: 1.13686367

(e) What is the relativity in the tariff for age category 4?

Answer: 0.78539832

(f) What is the expected claim frequency of a policyholder who has CONVT vehicle type, vehicle age 2, gender F, area D, and policyholder's age 3? Answer: 0.03641556 per unit of exposure.

(g) What is the expected claim frequency of a policyholder who has SEDAN vehicle type, vehicle age 1, gender F, area Others, and policyholder's age 1?

Answer: 1.13686367 per unit of exposure.

2 Chapter 9

2.1 Limited fluctuation credibility

1. How many losses are required for full credibility?

   Answer: 2,305

2. If the number of claims has a Negative Binomial distribution with $\beta = 1$, how many loses are required?

   Answer: 3,842

3. What is the full credibility standard for the pure premium (calculated as aggregate losses divided by exposures)?

   Answer: 2,305

2.2 Bühlmann credibility

1. For any risk in a population, the number of claims N in a year has a Poisson distribution with parameter λ. There are 2 classes of risks in the population, each with the same number of risks. The number of claims for each risk in Class A is a Poisson random variable with mean θ_A which has an exponential distribution with pdf $f(\theta_A) = e^{-\theta_A}$. The number of claims for each risk in Class B is a Poisson random variable with mean θ_B which has a Gamma (2,1) distribution. A risk was selected at random from the population and all claims were recorded over a five-year period. The total number of claims over the five-year period was 10.

   (a) What is the expected number of claims for an insured chosen at random from Class B?

   Answer: 2.

   (b) If a risk is selected at random from the population, what is the expected number of claims in a year?

   Answer: $0.5 \times 1 + 0.5 \times 2 = 1.5$.

   (c) It can be calculated that $\sigma^2(A) = 2$ and $\sigma^2(B) = 4$. Calculate the EPV and VMH for the population.

   Answer: EPV is $0.5 \times 2 + 0.5 \times 4 = 3$, VMH is $0.5 \times (1 - 1.5)^2 + 0.5 \times (1 - 2)^2 = 0.625$.

   (d) Use Bühlmann credibility to estimate the annual expected number of claims for the risk

Answer:

$$\begin{aligned}\bar{X} &= 10/5 = 2\\ K &= 3/0.625 = 4.8\\ Z &= 5/(5+4.8) = 0.51\\ \hat{\mu}(\theta) &= 0.51 * 2 + (1-0.51) * 1.5 = 1.755\end{aligned}$$

2.3 Bühlmann-Straub credibility

A commercial automobile policyholder had the following number of policies and total claim costs over a two-year period:

Year	Number of policies	Total claim cost
1	31	38,314
2	40	32,291

The total claim cost in a year for each policy in the policyholders fleet is Gamma distributed with shape parameter α and scale parameter 3. Parameter α is normal distributed with mean 300 and variance 2. The number of policies in year 3 is 43. Use Bühlmann-Straub credibility to estimate the expected total claim cost in year 3.

Answer:
Expected claim costs for one policy is 900.
Average cost per unit of exposure is 994.4366.
EPV is 2,700.
VHM is 18.
K is 150.
Number of exposures in the experience period is 71.
The credibility is 0.321267.
Credibility weighted estimated for the average claim cost for one unit of exposure is 930.3394.
Expected total cost for year 3 is 40,004.59.

2.4 Bayesian inference and Bühlmann credibility

The size of a claim X follows an Exponential distribution with density $f(x|\theta) = \theta e^{-\theta x}$. The parameter θ has a Gamma distribution with shape parameter α and rate parameterβ. The historical observations of claim sizes are $x_1, \ldots, x_n$.

1. Find the posterior distribution of θ.

   Answer: Gamma $(\alpha + n, \sum_{i=1}^{n} x_i + \beta)$.

2. Find the expression for the posterior expected claim size.

   Answer: $\dfrac{\sum_{i=1}^{n} x_i + \beta}{\alpha + n - 1}$

3. Show that the Bühlmann credibility estimate exact matches the Bayesian results.

   Answer:
   EPV is $\dfrac{\beta^2}{(\alpha-1)(\alpha-2)}$.
   VHM is $\dfrac{\beta^2}{(\alpha-1)^2(\alpha-2)}$.
   Credibility K is $\alpha - 1$.

Overall mean is $\dfrac{\beta}{\alpha-1}$.
Putting together, we can show that

$$\hat{\mu} = \frac{n}{n+\alpha-1}\frac{\sum x_i}{n} + \left(1 - \frac{n}{n+\alpha-1}\right)\frac{\beta}{n+\alpha-1} = \frac{\sum_{i=1}^{n} x_i + \beta}{\alpha+n-1}$$

2.5 Estimating credibility parameters

1. Two policyholders had total claims over a four-year period as shown in the table below. Estimate the expected total claims for each policyholder using Bühlmann credibility.

Year	1	2	3	4
Policyholder 1	230	265	198	240
Policyholder 2	156	150	140	180

Answer:

$$\hat{\mu}_A = 231.53 \qquad \hat{\mu}_B = 158.22$$

2. Three policyholders had claims shown in the table below. Calculate the nonparametric estimate for the VHM.

Policyholder		Year 1	Year 2	Year 3
A	Number of clams	0	2	1
A	Insured vehicles	1	2	2
B	Number of clams	0	0	1
B	Insured vehicles	0	2	3
C	Number of claims	1	2	1
C	Insured vehicles	2	4	3

Answer: It can be calculated that $VHM = -0.48$.
The process variance is so large that it is not possible to estimate the VHM.

3. (SOA exam, May 2010) The number of claims a driver has during the year is assumed to be Poisson distributed with an unknown mean that varies by driver. The experience for 100 drivers is as follows:

Number of claims during the year	Number of drivers
0	54
1	33
2	10
3	2
4	1

Determine the credibility of one year's experience for a single driver using Bühlmann credibility.

Answer: $EPV = 0.63, \quad VHM = 0.05, \quad Z = 0.073.$

3 Chapter 10

3.1 Tails of distributions

1. Let $X \sim$ Exponential(θ) with distribution function $f(x) = \lambda e^{\lambda x}$. Determine whether this distribution is light tailed or heavy tailed based on moments.
   Answer: $\mu'_k = \lambda^{-n} n! < \infty$, hence the distribution is light tailed.
2. Let $X \sim$ Gamma(α, θ), $\alpha > 1$, $\theta > 0$ and $Y \sim$ Exponential($\frac{1}{\theta}$). Compare the tail behaviour of these distributions using the limiting behaviour method.
   Answer: It can be easily shown that the gamma distribution is heavier tailed than the exponential distribution.

3.2 Risk measures

1. Check if the variance is a coherent risk measure.
   Answer: No. It can be shown that the variance is neither subadditive nor positively homogeneous.
2. Consider an insurance loss random variable $X \sim$ Pareto(α, θ), $\alpha > 0$, $\theta > 0$. Justify the closed form VaR expression $\theta[(1-q)^{-1/\alpha} - 1]$ of this distribution.
3. Which of the following is not correct about $TVaR$?
   A. It is the expected value of losses that exceed a specified VaR.
   B. TVaR of a strictly increasing linear transformation of a random variable is the function of TVaR of the original random variable.
   C. It is less sensitive to the choice of confidence level compared to VaR.
   D. **It is not subadditive.**
   E. It accounts for all events above the confidence level.
4. Calculate TVaR for variables X and Y in Example 10.2.5 to justify that Y is riskier than X.
   Answer: TVaR of X is 97.5, TVaR of Y is 126.7047.

3.3 Reinsurance

1. Summarise some basic reinsurance treaties by proportional/non proportional categories and expressions for the amounts paid by direct insurer and reinsurer in the following table:

Type	Prop/Non-prop?	Total loss	Paid by insurer	Paid by reinsurer
Quota share (c)		X		
Stop-loss (M)		X		
Excess of loss (M_i)		$X = \sum_{i=1}^{n} X_i$		

Answer:

Type	Prop/Non-prop?	Total loss	Paid by insurer	Paid by reinsurer
Quota share (c)	Proportional	X	cX	$(1-c)X$
Stop-loss (M)	Non-proportional	X	$\min(X, M)$	$\max(0, X-M)$
Excess of loss (M_i)	Non-proportional	$X = \sum_{i=1}^{n} X_i$	$\sum_{i=1}^{n} \min(X_i, M_i)$	$X - \sum_{i=1}^{n} \min(X_i, M_i)$

2. Consider the graph in Example 10.3.2 and answer the following questions

   (a) Why do the values c_1, c_2, c_3 increase linearly with K?

   (b) Why does the order of lines $c_1 > c_2 > c_3$ from top to bottom of the graph?

   Answer:

   (a) We have $K = \sum_{i=1}^{3} c_i E(X_i)$, hence K is a linear function of c_i.

   (b) The behaviour of the lines $c_1 > c_2 > c_3$ is a result of the distributions they apply to. We have that $c_i \propto \frac{E(X_i)}{Var(X_i)}$. Other things being equal, a higher revenue as measured by $E(X_i)$ means a higher value of c_i. In the same way, a higher value of uncertainty as measured by $Var(X_i)$ means a lower value of c_i. We have the following mean and variance of X_1, X_2, X_3:

      - $E(X_1) = 500$; $Var(X_1) = 1e + 06 - 500$
      - $E(X_2) = 1000$; $Var(X_1) = 4e + 06 - 1000$
      - $E(X_2) = 1000$; $Var(X_1) = 3e + 06 - 1000$

      It can then be seen easily that $c_1 > c_2 > c_3$.

3. Consider a surplus share proportional treaty where the retained line is 10,000 and the limit is 50,000. Let X be the total loss. Write down the expression for the loss paid by the insurer and the reinsurer and draw diagrams of their payments (on the y axis) with respect to the total loss (on the x axis).
   Answer: Insurer pays

$$\begin{cases} X, & \text{if } X \leq 10,000, \\ 10,000, & \text{if } 10,000 < X \leq 60,000, \\ X - 50,000, & \text{if } X > 60,000. \end{cases}$$

   Reinsurer pays

$$\begin{cases} 0, & \text{if } X \leq 10,000, \\ X - 10,000, & \text{if } 10,000 < X \leq 60,000, \\ 50,000, & \text{if } X > 60,000. \end{cases}$$

   Diagrams can then be drawn using these expressions.

4. For the total loss and each of the parties in Example 10.3.4, draw a diagram of their payments (on the y axis) with respect to the total loss (on the x axis).

5. Consider example 10.3.6. Now you also consider retaining 30% of the directors and executive officers risk X_3 and 20% of the cyber risks X_4, in addition to existing stop loss arrangement of building risks X_1 and motor vehicles risks X_2. Answer questions (a)-(c) in Example 10.3.6 for this new portfolio. How does the retained portfolio risk in this case compared to when X_3 and X_4 are fully covered by insurer?
   Answer:

   (a) Expected claim amount

Retained	Insurer	Total
752.41	1791.05	2543.46

   (b) 80th, 90th, 95th, and 99th percentiles

	80%	90%	95%	99%
Retained	937.55	1284.00	1713.35	3188.34
Insurer	2418.87	3424.39	4745.83	9084.53
Total	3351.35	4675.04	6464.20	12159.02

(c) Since the retained risk now includes parts of X_3 and X_4, the portfolio risk is longer tailed compared to when these risks are not included.

4 Chapter 13

4.1 Some R functions

Consider the dataset "Anscombe" in the R package "carData". This data contains US state public-school expenditures.

1. Provide descriptive statistics of the variables "education", "income", "young" and "urban" in this dataset.
   Answer:

education	income	young	urban
Min. :112.0	Min. :2081	Min. :326.2	Min. : 322.0
1st Qu.:165.0	1st Qu.:2786	1st Qu.:342.1	1st Qu.: 552.5
Median :192.0	Median :3257	Median :354.1	Median : 664.0
Mean :196.3	Mean :3225	Mean :358.9	Mean : 664.5
3rd Qu.:228.5	3rd Qu.:3612	3rd Qu.:369.1	3rd Qu.: 790.5
Max. :372.0	Max. :4425	Max. :439.7	Max. :1000.0

2. Perform k-means cluster analysis with 5 clusters on this dataset. Note that you need to scale the data set first (using function "scale(.)") to standardise explanatory variable units. What are the size of these clusters?
   Answer: 1, 16, 8, 15, 11
3. Plot a histogram of the "education" variable. Fit a gamma distribution to this variable.
   Answer: Parameter estimates: 19.31636309, 0.09839537.
4. Perform a linear regression with "education" variable as the explanatory variable and "income" as the response variable.
   Answer: Coefficients: 1645.383 (intercept), 8.048.

5 Chapter 14

5.1 Variable types

1. Identify the type of variable for each of the following:
   (a) Satisfactory level on the scale from 1 to 5 in a teaching evaluation survey. (Answer: ordinal variable)
   (b) The breed of a dog. (Answer: nominal variable)
   (c) Answer to a True or False question. (Answer: binary variable)

(d) Number of people living in New York city. (Answer: discrete variable)

(e) Gas price. (Answer: continuous variable)

5.2 Classic measures of scalar association

1. In this question you will be asked to perform some analyses on the "dataCar" data set available in the R package "insuranceData". You first need to install the package (if you do not have it installed yet), and load the dataset.

   (a) Calculate the Pearson correlation, Spearman's rho and Kendall's tau correlation between "exposure" (exposure) and "numclaims" (number of claims) variables.
   Answer: Pearson: 0.1344, Spearman's rho: 0.1337, Kendall's tau: 0.1091

   (b) Without doing any calculations, do you think these correlations will change if we:

      i. scale the variable "numclaims" with a multiplication factor of 2.

      ii. apply a log transformation to the variable "exposure".

   Confirm your answer by performing the calculations in R.
   Answer: i. No, ii. Pearson correlation changes, Spearman's rho and Kendall's tau are unaffected.

   (c) Calculate the odds ratio between the two variables "clm" (claims) and "gender" (gender). How do you interpret this result?
   Answer: 0.9836

   (d) Test whether the variable "clm" is independent of the variable "area " (area) using Pearson chi-square statistics. Do you get the same answer when you use the likelihood ratio test?
   Answer: Both tests have p-value less than the 0.05 significance level, we reject the null hypothesis that these variables are independent.

   (e) Which type of correlation do you use to assess the relationship between "clm" and "agecat" (age category) variables? Calculate this correlation in R. What is the nature of their relationship?
   Answer: Polychoric correlation = -0.06. There is a negative dependence between the two variables.

5.3 Application using copulas

Consider the "loss" data set studied in this section.

1. Fit log-normal distributions to the "expenses" and "losses" variable. You may consider using the "fitdistr" from the R MASS package.
   Answer: "losses" parameter estimates: 9.37345394 1.63756011, "expenses" parameter estimates: 8.52197632 1.42942232.

2. Perform probability integral transformation on these variables with log-normal fitting. Plot the histogram of the transformed variables.

3. Transform the two variables "expenses" and "losses" to normal scores and plot the histograms of these scores.

4. Draw scatter plots of uniform transformed variables and of normal scores obtained in questions 2 and 3. Calculate the Spearman's rho correlation and compare the result with the Spearman's rho correlation calculated with Pareto fitting in the illustration.

Answer: Spearman's rho: 0.4519872 which is the same as the Spearman's rho with Pareto fitting. (Do you know why?)

5. Fit Frank's copula with the maximum likelihood method to the transformed uniform variables. Compute the Spearman's correlation corresponding to the fitted copula parameter. Compare the copula estimates and Spearman's correlation with the results from Pareto fitting in the illustration.
   Answer: Frank's copula parameter estimate: 3.176 and Spearman's rho: 0.469. These are not the same as the copula and correlation with Pareto fitting.

6. Another type of bivariate copula is Gumbel-Hougaarg specified by

$$C_\theta^{GH}(u) = \exp\left(-\left(\sum_{i=1}^{2}(-\log u_i)^\theta\right)^{1/\theta}\right), \quad \theta \in [1, \theta)$$

   Fit this copula with maximum likelihood method to the transformed uniform variables. Compute the Spearman's correlation corresponding to the fitted copula parameter and compare it with the Spearman's rho correlation from the Frank's copula.
   Answer: Gumbel's parameter estimate: 1.457 and Spearman's rho: 0.449. The Spearman's rho is not the same as the one implied from the previously fitted Frank's copula.

7. What conclusion can you draw regarding the uniqueness of copula on marginal choices from this exercise?
   Answer: The estimate of copula depends on the choice of marginal. Spearman's correlation also varies depending on the type of copula chosen. However, it does not depend on the choice of marginals.

5.4 Types of copulas

1. Verify the elliptical copula generator for normal distribution, t-distribution and Cauchy distribution in table 14.6.

2. Find the generator functions of Frank copula, Clayton copula and Gumbel-Hougaard copula respectively (for simplification, you can consider the bivariate case).
   Answer:
   Frank copula:

$$g(x) = -\log\left(\frac{e^{\theta x} - 1}{e^\theta - 1}\right)$$

   Clayton copula:

$$g(x) = \frac{1}{\theta}\left(x^{-\theta} - 1\right)$$

   Gumbel-Hougaard copula:

$$g(x) = (-\log(x))^\theta$$

5.5 Why is dependence modeling important?

1. [AV - If R code for simulation using Gaussian copula is made available] Obtain expected value and quantiles for the portfolio in the example in this chapter using t copula. Comment on impact of the choice of copula on quantiles of the portfolio.