Eva Kotlebová,
Ľudovít Pinda

The Important Role of Bayesian Statistics in the Decision Making Process and Management of Insurance Companies

Article Info:

Management Information Systems, Vol. 8 (2013), No. 2, pp. 003-007

Received 21 March 2013
Accepted 24 April 2013

UDC 519.226:368.03

Summary

The article deals with the Bayesian statistics, which is appropriate in situations when there is lack of available statistical data, so that another source of information must be included into the analysis. The main mathematical principle of Bayesian approach is described. We focused on the analysis of variance, which is usually (in classical statistics) used for deciding whether the factor influences the quantitative variable. Besides this decision, The Bayesian approach also provides the possibility of estimating the group's means, different from the classical approach. Such kind of estimation (Bayesian shrinkage point estimation) is more precise, and therefore more valuable for consequential analyses and decisions. Processing real data of car insurance, the rate of influence of engine performance for the claim amount was shown and the group's means were estimated. Such estimation may be used for suggesting the insurance rate in each group for the next period.

Keywords

Bayesian approach, prior distribution, posterior distribution, conjugate family, analysis of variance, shrinkage point estimation, insurance rate

1. Introduction

Any responsible decision must be based on thorough knowledge of the problem that is being solved. The knowledge usually comes from long-term experience, but it has to be supported by indubitable facts, too. Statistics provides analyses based on processing real data. Such analyses are noted for high measure of objectiveness, and thus constitute a valuable background for making well-advised decisions.

Sometimes it is a great problem to gather enough data for describing the whole population. Methods of statistical inference are used in such a situation, as they allow discovering some substantial information from the sample, where the units are randomly gathered from the population. The main problems statistical inference deals with are estimations of the population's parameters (mean, variance, proportion, etc.) and testing hypotheses about some properties of the population.

Bayesian statistics is an effective tool for solving some inference problems in a situation when the available sample is too small for more complex statistical analysis to be applied. The lack of information may be offset (up to a certain point) by using Bayesian approach, as it enables us to utilise more sources of information. Besides the sample data, so-called prior information may be included into the analysis.

Bayesian approach is very well applicable in insurance. In order to set the optimal insurance rate, it is necessary to know the distribution of several variables: number of insured events, the volume of insurance cover and the gross amount of insurance cover. The only way to create these distributions is to use available data coming from the previous periods. The more data is employed, the better the estimated distribution is. Thus, any other source of data is desirable to utilize. Bayesian statistics shows how to include both kinds of information to the analysis.

2. The Main Principle of Bayesian Statistics

The main mathematical difference between the Bayesian and the classical approach in solving inference problems is that while in the classical statistical inference the estimated parameter is an unknown constant, in Bayesian statistics the parameter is considered to be a random variable with some kind of distribution. Before processing the data from random sample, the distribution is called prior distribution; incorporating the data leads to the posterior distribution, which is the background for making inference conclusions – estimations of parameters and testing hypotheses.

The main principle, Bayesian statistics is based on is the Bayes' theorem (Bolstad, 2004), which expresses how the occurrence of a particular event influences the probabilities of the other events. Its discrete form is as follows:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{\sum_{j=1}^{n} P(B \mid A_j) \cdot P(A_j)} \quad (1)$$

($A_1, A_2, ..., A_n$ is a sequence of exclusive and exhaustive events, is any event.)

The rules of Bayesian statistics are mostly derived from the continuous form:

$$f_\Theta(\theta \mid \mathbf{x}) = \frac{f(\mathbf{x} \mid \theta) \cdot f_\Theta(\theta)}{\int_\Theta f(\mathbf{x} \mid \theta) \cdot f_\Theta(\theta)\, \mathbf{d}\theta} \quad (2)$$

where

$f_\Theta(\theta)$ denotes the prior probability of estimated parameter Θ

$f_\Theta(\theta \mid \mathbf{x})$ is the posterior density of Θ

$f(\mathbf{x} \mid \theta)$ is the likelihood function.

The expression in denominator can be neglected, because considering the variable θ, it is of constant value. Thus, the relation of identity is substituted by the relation of proportionality and the theorem is mostly used in following form:

$$f_\Theta(\theta \mid \mathbf{x}) \propto f(\mathbf{x} \mid \theta) \cdot f_\Theta(\theta) \quad (3)$$

So, the posterior density depends on the sample distribution as well as the prior density. The rate of these influences is given by several factors; the most important of them are related to some properties of the sample, i.e. its size and its variance.

Generally, the more valuable source is that of the less variance, so the smaller the sample variance, the bigger the influence of the sample distribution over the posterior distribution's shape, and vice versa.

It is obvious that the influence of sample distribution rises if the sample size increases. The variance of the posterior distribution is usually smaller than the prior variance (and the sample variance too); thus the conclusions based on posterior distribution are more precise than those provided by means of the classical approach.

Sometimes, the prior and the posterior densities are of the same type, although the values of their parameters are different. In such cases, together with the sample distribution, they create so-called conjugate family. Employing conjugate families makes considerably easier to evaluate the parameters of posterior distribution, as there exist formulas for doing it.

The most used conjugated families in insurance are these (the first is the sample distribution, the second is the prior/posterior distribution):

<u>Binomial/beta</u> (Lee, 2012; Pacáková, 2004; Kotlebová, 2009) is used for estimating the probability of the insurance event in the case when we consider this probability to be constant. The sample comes from the binomial distribution $Bi(n;\pi)$, and the prior distribution of π is beta distribution $Be(\alpha;\beta)$. The posterior distribution of π is beta distribution, too, with parameters $\alpha_1 = \alpha + x$, $\beta_1 = \beta + n - x$. (x denotes number of the event's occurrences.)

<u>Poisson/gamma</u> (Lee, 2012; Pacáková, 2004; Kotlebová, 2009) is used for estimating the distribution of a number of insured events. The sample comes from the Poisson distribution $Po(\lambda)$, and the prior distribution of λ is gamma distribution $G(\alpha;\beta)$. The posterior distribution is gamma distribution $G(\alpha_1;\beta_1)$; the parameters satisfy equations: $\alpha_1 = \alpha + \sum_{i=1}^{n} x_i$, $\beta_1 = \beta + n$. (n is number of observations, x_i are observed values).

<u>Normal/normal</u> (Lee, 2012; Pacáková, 2004; Kotlebová, 2009) is used for estimating the average volume of insurance cover.

The sample is normally distributed, and the variance of the distribution is assumed to be known (σ^2). If the prior distribution of mean (μ) is normal ($N(m;s^2)$), the posterior is normal, too ($N(m_1;s_1^2)$); the parameters satisfy equations:

$$m_1 = \frac{\frac{1}{s^2} \cdot m + \frac{n}{\sigma^2} \cdot \bar{x}}{\frac{1}{s^2} + \frac{n}{\sigma^2}}, \qquad s_1^2 = \frac{1}{\frac{1}{s^2} + \frac{n}{\sigma^2}}$$

Usually, the point estimation of the parameter is the posterior mean, the limits of confidence intervals are the appropriate quantiles of the posterior distribution.

3. Bayesian Approach in Analysis of Variance

In this article, we concentrated on the Bayesian approach in one of the most often used methods in statistics - analysis of variance (ANOVA). To briefly remind the main principle of this method:

The aim is to decide, whether some factor (character variable) influences the particular quantitative variable. In order to do that, the data is selected into groups – each of them is created by particular variant of the factor. The goal is to verify hypothesis:

$$H_0 : \mu_1 = \mu_2 = ... \mu_k \ \ (k > 2), \qquad (4)$$

where $\mu_1, \mu_2, ..., \mu_k$ are groups' means and k is the number of groups (variants of factor). If the hypothesis holds, the factor does not influence the quantitative variable. If any of equations is not satisfied, we consider the factor to be statistically significant.

The following F-test is used for verifying the hypothesis (4):

$$F = \frac{MSA}{MSE} = \frac{\frac{SSA}{k-1}}{\frac{SSE}{n-k}} \qquad (5)$$

with $(k-1)$ and $(n-k)$ degrees of freedom. It is based on comparing two kinds of variance – variance between groups and within groups. If the calculated value is bigger than the quantile $F_{1-\alpha(k-1;n-k)}$, we reject the null hypothesis and we conclude, that the factor influences the value of the quantitative variable.

In the following part, we concentrate on estimating means of the particular group μ_i. The classical statistics approaches this problem in a simple way: When the null hypothesis is rejected, the group's mean is estimated by average of values belonging to the group $\bar{x}_i$. If the null hypothesis is not rejected, means of each group are estimated by the total mean $\bar{\bar{x}}$ (exploitation more data gives better result). So, the classical approach leads to two different estimations. It is obvious, that the result depends on the significance level, too.

Bayesian statistics provides more possibilities for the estimated group´s mean. If the factor appears to be statistically significant, besides the decision about the hypothesis (4) Bayesian statistics also takes into account the difference between the evaluated F-statistics and the quantile $F_{1-\alpha(k-1;n-k)}$. The reason for such consideration is as follows:

Sometimes, the evaluated F-statistics hardly exceeds the quantile $F_{1-\alpha(k-1;n-k)}$. The difference may be such small, that some doubts about the decision (of rejecting the null hypothesis) may appear. If, in addition, the difference between the over-all average and the group´s average is sizeable, we can recognize, that tiny change in data leading to decrease the F-statistics value, causes the big change in estimation of the mean´s value. Thus, the question is, whether the better estimation of the group´s mean is somewhere within both averages under consideration. Bayesian statistics gives a reasonable answer to this question. The group´s mean may be estimated by using the formula

$$BE(\mu_i) = \left(\frac{1}{F}\right) \cdot \bar{\bar{x}} + \left(1 - \frac{1}{F}\right) \cdot \bar{x}_i \qquad (6)$$

Accordingly, Bayesian estimation of the group´s mean $BE(\mu_i)$ is a weighted average of both averages. The weights appropriately represent the relations listed above. The greater the value of F-statistics (which lies in the critical region far from the critical value, so the decision about rejection the null hypothesis is unambiguous), the smaller the weight of the total average, and vice versa. This kind of estimation is called *Bayesian shrinkage estimation*, as the groups' mean is „shrinkaged" towards the total mean.

The formula may be used without reference to the significance level used for testing the hypothesis (4).

4. Bayesian Approach in Estimation of Average Claim Amount Depending on Engine Performance of a Car

We used the described procedure on the data coming from an insurance company. The authentic data were transformed, according to the requirement of the insurance company which provided the data. (That is the reason, why we don´t record money units in Table 2 and Table 3.)

Tables 1 and Table 2 contain data about number of insurance contracts and the aggregated claim amount. They have been sorted according to engine performance, and the clients were legal persons. Symbols B1, B2, …, B6 represent the following values of engine performance:

B1 – engine performance up to 50 kW
B2 – engine performance 51 – 60 kW
B3 – engine performance 61 – 80 kW
B4 – engine performance 81 – 90 kW

B5 – engine performance 91 – 120 kW
B6 – engine performance greater than 120 kW

Table 1 Number of insurance contracts (transformed)

Number of insurance contracts						
Year	B1	B2	B3	B4	B5	B6
2007	6 672	6 479	9 197	1 562	1 564	1 045
2008	7 546	6 376	7 941	1 303	1 242	1 054
2009	7 536	6 458	9 728	2 124	2 130	1 923
2010	12 021	10 521	26 132	6 612	9 335	6 650
2011	12 191	10 885	29 300	7 212	10 218	7 689
Total	45 966	40 719	82 298	18 813	24 489	18 361

Source: Insurance company

Table 2 Aggregate claim amount (transformed)

Aggregate claim amount						
Year	B1	B2	B3	B4	B5	B6
2007	350 011	703 163	620 861	113 207	225 912	112 382
2008	319 812	395 520	395 555	102 598	122 409	62 091
2009	265 313	238 976	435 837	61 571	101 971	112 669
2010	470 582	540 519	1 382 465	324 763	616 843	484 000
2011	428 357	355 235	1 242 557	319 656	730 292	444 466
Total	1 843 075	2 233 413	4 077 275	921 795	1 797 427	1 215 608

Source: Insurance company

Using data from Tables 1 and 2 we evaluated the average claim amount related to one insurance contract in each group. The results are presented in Table 3.

Table 3 Average claim amount related to one insurance contract

Average claim amount related to one insurance contract						
Year	B1	B2	B3	B4	B5	B6
2007	52.459682	108.529557	67.506044	72.475672	144.445013	107.542584
2008	42.381659	62.032622	49.811737	78.739831	98.557971	58.909868
2009	35.206077	37.004645	44.802323	28.988230	47.873709	58.590224
2010	39.146660	51.375250	52.903146	49.117211	66.078522	72.781955
2011	35.137150	32.635278	42.408089	44.322795	71.471129	57.805436

Source: The authors' calculations from transformed data

We wanted to find out whether the engine performance influences the average claim amount, and estimate the groups´ means (regardless of the tests' conclusion). The Statgraphics Plus software was used for the calculations. The first output is in Figure 1.

Summary Statistics for claim amount

engine performance	Count	Average	Variance
B1	5	40,8662	51,1303
B2	5	58,3155	924,265
B3	5	51,4864	97,1402
B4	5	54,7287	423,458
B5	5	85,6853	1408,65
B6	5	71,126	453,182
Total	30	60,368	678,995

Figure 1 Output from Statgraphics Plus (Summary statistics)

The first sight shows, that the means are considerably higher in the last two groups in comparison with other groups. However, the variability within groups seems great, too, so it is hard to decide in advance whether the factor is statistically significant.

Before performing analysis of variance, we had to verify whether the variances in groups are the same or not (Pacáková, 2003). It turned out, that the p-Value in the Cochran´s test was 0.135418, so the necessary condition for performing ANOVA was satisfied.

The ANOVA table is shown in Figure 2, whereas the particular groups are graphically compared in Figure 3.

ANOVA Table for claim amount by engine performance
Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Between groups	6259,57	5	1251,91	2,24	0,0834
Within groups	13431,3	24	559,637		
Total (Corr.)	19690,9	29			

Figure 2 Output from Statgraphics Plus (ANOVA table)

According to the p-Value, the null hypothesis is rejected at the significance level 0.05, while it is not rejected at the significance level 0.1. Thus, according to the classical approach, at the significance level 0.05, all groups' means are estimated by total average 60.368, while at the significance level 0.1 each mean is estimated by the group´s average (Figure 1). The differences are sizeable, particularly in groups B5 and B1. (The Kruskall-Wallis (Pacáková, 2003) test was performed as well; p-Value was also within 0.05 and 0.1, so the estimations are the same).

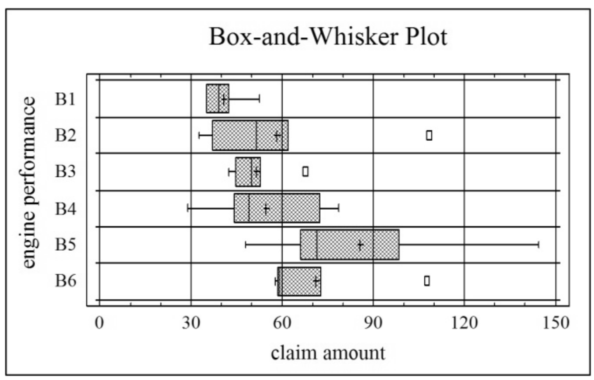

Figure 3 Output from Statgraphics Plus

By the Bayesian approach, for the groups' means estimation, we used formula (6), to which the value ($F = 2.24$) from Table 2 was substituted. Calculations are in Table 4. The values of used weights were these:

$$\frac{1}{F} = \frac{1}{2.24} = 0.446429, \; 1 - \frac{1}{F} = 0.553571$$

Table 4 Calculated Bayesian shrinkage point estimation

Engine performance	Classical point estimation at significance level 0.1	Classical point estimation at significance level 0.05	Bayesian shrinkage estimation
B1	40.8662	60.638	49.5724
B2	58.3155	60.638	59.2318
B3	51.4864	60.638	55.4514
B4	54.7287	60.638	57.2462
B5	85.6853	60.638	74.3829
B6	71.1260	60.638	66.3233
Total	60.3680	60.638	60.3680

Source: The authors' calculations

Thus, each Bayesian shrinkage estimation is within the groups' average and the total mean. In Bayesian statistics, generally, point estimations are within the classical point estimation (coming from the sample data) and the value derived from the prior information. In this case, the prior assumption was connected with the null hypothesis (4), thus the derived value was the total mean. Calculated values may create a credible base for making decisions about the insurance premium rate.

5. Conclusion

Well-advised decisions are based on arguments derived from the real data. If there is too little data at disposal, it is desirable to employ another source of information. Bayesian statistics provides methods which pool several kinds of information together. The achieved results are more comprehensive than those coming from the classical approach.

Bayesian approach in analysis of variance allows us not only to decide whether the variable is influenced by the factor, but it provides also a simple tool for estimating the groups' means, different from the classical approach. Bayesian shrinkage point estimation gives valuable results especially in situations, when the p-Value of F-test (when performing ANOVA) is not far from the significance level.

As it was shown, the estimations provided by the classical approach were considerably different for two different significance levels (0.05, 0.1) (when performing the ANOVA test); Bayesian estimation is a suitable middle course within both classical possibilities.

References

Anderson, A. (2012). *Understand the Math Behind it All: Bayesian Statistics.* Retrieved January 15, 2013, from Adobe Blogs: http://blogs.adobe.com/digitalmarketing/personalization/conversion-optimization/understand-the-math-behind-it-all-bayesian-

Bernardo, J. M., & Smith, A. F. (1995). *Bayesian Theory.* Hoboken: John Wiley & Sons.

Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F., et al. (2011). *Bayesian Statistics 9 (Eds.).* Oxford: Oxford University Press.

Bolstad, W. M. (2004). *Introduction to Bayesian Statistics.* Hoboken: John Wiley & Sons.

Christensen, R., Johnson, W., Branscum, A., & Hanson, T. E. (2011). *Bayesian Ideas and Data Analysis.* Boca Raton: CRC Press ,Taylor& Francis Group.

Freund, J. E. (1992). *Mathematical statistics.* Upper Saddle River: Prentice-Hall International.

Garthwaite, P. H., Jolliffe, I., & Jones, B. (1995). *Statistical Inference.* Upper Saddle River: Prentice-Hall International.

Horáková, G., & Huťka, V. (2002). *Teória pravdepodobnosti 1.* Bratislava: Vyd. EKONÓM.

Kotlebová, E. (2009). *Bayesovská štatistická indukcia v ekonomických aplikáciách.* Bratislava: Vyd. EKONÓM.

Lee, P. (2012). *Bayesian statistics.* Hoboken: John Wiley & Sons.

Pacáková, V. (2004). *Aplikovaná poistná štatistika.* Bratislava: IURA EDITION.

Pacáková, V. (2012). *Štatistická indukcia pre ekonómov.* Bratislava: Ekonóm.

Pacáková, V. (2003). *Štatistika pre ekonómov.* Bratislava: IURA EDITION.

Robert, P. C. (2006). *Bayesian Statistics.* Paris: Université Paris Daphine and CREST-INSEE.

Šoltés, E. (2009). *Modely kredibility na výpočet poistného.* Bratislava: Vyd. EKONÓM.

Terek, M. (2003). *Úvod do analýzy rozhodovania a bayesovskej indukcie.* Bratislava: Vyd. EKONÓM.

Wonnacott, T. H., & Wonnacott, R. J. (1993). *Statistika pro obchod a hospodářství.* Prague: Victoria Publishing.

Eva Kotlebová
University of Economics, Bratislava
Department of Statistics, Faculty of Business Informatics
Dolnozemská cesta 1
852 35 Bratislava
Slovakia
Email: eva.kotlebova@gmail.com

Ľudovít Pinda
University of Economics, Bratislava
Department of Mathematics, Faculty of Business Informatics
Dolnozemská cesta 1
852 35 Bratislava
Slovakia
Email: ludovit.pinda@euba.sk