QUADRATIC QUASINORM
AND ITS APPLICATIONS IN RISK ANALYSIS

Jakub Šácha[1], Veronika Lacinová[2], Zdeněk Karpíšek[3]

[1] Department of Statistics and Operation Analysis, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

[2] Department of Quantitative Methods, Faculty of Military Leadership, University of Defence, Kounicova 65, 662 10 Brno, Czech Republic

[3] Department of Applied Disciplines, STING Academy, Stromovka 1, 637 00 Brno, Czech Republic

Abstract

This paper deals with the estimation of the probability distribution of the category random variable from its observed values. The gradient estimation presented is based on the f-quasi-norm term. The gradient estimation tends to move away from the empirical distribution towards a uniform distribution. For this reason, it is called the pessimistic estimate. The paper deals with both point and interval estimates. The bootstrap method is used for interval estimates. The theoretical results are illustrated using specific economic data.

Keywords: f-divergence, f-quasi-norm, quadratic quasi-norm, discrete probability distribution, gradient estimation, bootstrap

INTRODUCTION

The estimation of probability distribution from the observed random sample x_i, $i = 1, ..., n$, where $n > m$ is a basic practical task in stochastic modeling of a categorical random variable X which takes a finite number of different values x_j^*, $j = 1, ..., m$, where $m \geq 2$.

A different gradient estimation $p(t)$ of a discrete random variable is presented in this paper, which can be called pessimistic in a certain sense. The estimate is based on the concept of f-divergence D_f (Vajda, 2004), which is used to measure the "distance" of two discrete distributions within the same dimension. Using the f-divergence term, the f-quasi-norm of distribution p is introduced as a "distance" $D_f(p, p_0)$ from the uniform distribution p_0 (Karpíšek *et al.*, 2007). The uniform distribution was chosen because it has the maximum uncertainty in a given probability space, for example in the sense of Shannon entropy.

If we move along the trajectory of the greatest slope of the used f-quasinorm from the observed distribution to the uniform distribution, the gradient estimate is constructed as the distribution nearest to the even distribution, while the Goodness of Fit Test does not yet reject null hypothesis that there is no significant difference between the observed and the estimated distribution. The problem leads to the solution of a system of ordinary differential equations of the first order, which are constructed in this paper for the quadratic f-quasinorm for which this system is linear. The bootstrap method is used to improve gradient estimates for small data sets and to construct interval estimates.

MATERIALS AND METHODS

In what follows we use the following concepts (Karpíšek *et al.*, 2007; Karpíšek *et al.*, 2008; Karpíšek and Neradová, 2014).

Let $S = \left\{ \boldsymbol{p} \in R^m : \forall p_j \geq 0, \sum_{j=1}^{m} p_j = 1 \right\}$ be the set of all discrete probability distributions on Ω. We define a **quasi-norm** of $\boldsymbol{p} \in S$ as the *f*-divergence $D_f(\boldsymbol{p}, \boldsymbol{p}_0)$ where

$$\boldsymbol{p}_0 = \left(\frac{1}{m}, \ldots, \frac{1}{m} \right)$$

and say that the function *f* generates the quasi-norm $D_f(\boldsymbol{p}, \boldsymbol{p}_0)$ on S.

This article focuses on the quadratic f-quasinorm, which we obtain from χ^2-divergence $\chi^2(\boldsymbol{p}, \boldsymbol{q})$ which is generated by the function $f(u) = (u - 1)^2$ (Karpíšek *et al.*, 2008)

$$D_{(u-1)^2}(\boldsymbol{p}, \boldsymbol{p}_0) = K(\boldsymbol{p}, \boldsymbol{p}_0) = \frac{1}{m} \sum_{j=1}^{m} (m p_j - 1)^2.$$

The idea of estimating the distribution $\boldsymbol{p}$ from observed values of variable *X* is based on finding a distribution in *S* that is the closest to

$$\boldsymbol{p}_0 = \left(\frac{1}{m}, \ldots, \frac{1}{m} \right),$$

which can be arrived at from the empirical distribution

$$\left(\frac{f_1}{n}, \ldots, \frac{f_m}{n} \right)$$

by a procedure which, in a certain sense is the fastest. This can be done by suitably minimizing a chosen quasi-norm $D_f(\boldsymbol{p}, \boldsymbol{p}_0)$ and looking for the distribution $\boldsymbol{p}$ on the steepest-descent contour in *S*.

As the gradient estimate of a probability distribution $\boldsymbol{p} \in S$ based on a sample distribution

$$\left(\frac{f_1}{n}, \ldots, \frac{f_m}{n} \right) \neq \left(\frac{1}{m}, \ldots, \frac{1}{m} \right)$$

we take to be such a probability distribution $\boldsymbol{p}(t) \in S$ that

$$\frac{d}{dt} \boldsymbol{p}(t) = -grad D_f(\boldsymbol{p}(t), \boldsymbol{p}_0) \text{ for } \forall t \in [0; \infty),$$

$$\boldsymbol{p}(0) = \frac{\boldsymbol{f}}{n} = \left(\frac{f_1}{n}, \ldots, \frac{f_m}{n} \right).$$

If a function $f(u)$ which generates quasi-norm $D_f(\boldsymbol{p}, \boldsymbol{p}_0)$ to *S*, has the necessary properties for the generating function and has a continuous derivative $f'(u)$ for $\forall u \in (0; \infty)$, then there is a gradient estimate $\boldsymbol{p}(t) = (p_1(t), \ldots, p_m(t))$ where $\forall t \in [0; \infty)$ of the probability distribution $\boldsymbol{p} \in S$. Its components $p_1(t), \ldots, p_{m-1}(t)$ are a particular solution of the system of first-order ordinary differential equations of the first order, which is non-linear in general and whose solution requires application of

a numerical method. The parameter *t* is understood as a parameter of the solution curve. An exception for which an explicit solution can be found is the quadratic quasi-norm for which the components of the gradient estimate $\boldsymbol{p}(t) = (p_1(t), \ldots, p_m(t))$ from the sample distribution

$$\left(\frac{f_1}{n}, \frac{f_2}{n}, \ldots, \frac{f_m}{n} \right)$$

are, for $\forall t \in [0; \infty)$ a particular solution of the non-homogeneous linear system of first-order ordinary differential equations (LSODE1) with constant coefficients and constant right-hand sides

$$p_1'(t) = -4m p_1(t) - 2m p_2(t) - \ldots - 2m p_{m-1}(t) + 2m,$$

$$p_2'(t) = -2m p_1(t) - 4m p_2(t) - \ldots - 2m p_{m-1}(t) + 2m,$$

$$\ldots,$$

$$p_{m-1}'(t) = -2m p_1(t) - 2m p_2(t) - \ldots - 4m p_{m-1}(t) + 2m,$$

with the initial conditions

$$p_1(0) = \frac{f_1}{n}, p_2(0) = \frac{f_2}{n}, \ldots, p_{m-1}(0) = \frac{f_{m-1}}{n}$$

and

$$p_m(t) = 1 - \sum_{j=1}^{m-1} p_j(t) \text{ for } \forall t \in [0; \infty).$$

The gradient estimate $\boldsymbol{p}(t)$ has the components

$$p_1(t) = c_1 e^{-2m^2 t} + c_2 e^{-2mt} + \frac{1}{m},$$

$$p_2(t) = c_1 e^{-2m^2 t} + c_3 e^{-2mt} + \frac{1}{m},$$

$$\ldots,$$

$$p_{m-2}(t) = c_1 e^{-2m^2 t} + c_{m-1} e^{-2mt} + \frac{1}{m},$$

$$p_{m-1}(t) = c_1 e^{-2m^2 t} - c_2 e^{-2mt} \ldots - c_{m-1} e^{-2mt} + \frac{1}{m},$$

$$p_m(t) = -(m-1) c_1 e^{-2m^2 t} + \frac{1}{m},$$

where

$$c_1 = \frac{\frac{f_1}{n} + \frac{f_2}{n} + \cdots + \frac{f_{m-1}}{n}}{m-1} - \frac{1}{m},$$

$$c_2 = \frac{\frac{(m-2)f_1}{n} - \frac{f_2}{n} - \cdots - \frac{f_{m-2}}{n} - \frac{f_{m-1}}{n}}{m-1},$$

$$c_3 = \frac{-\dfrac{f_1}{n} + \dfrac{(m-2)f_2}{n} - \cdots - \dfrac{f_{m-2}}{n} - \dfrac{f_{m-1}}{n}}{m-1},$$

$$\ldots$$

$$c_{m-1} = \frac{-\dfrac{f_1}{n} - \dfrac{f_2}{n} - \cdots + \dfrac{(m-2)f_{m-2}}{n} - \dfrac{f_{m-1}}{n}}{m-1}.$$

The gradient estimate so obtained is an estimate with a single parameter of the probability distribution. A suitable value t_0 can be found by a goodness-of-fit test. When using the Pearson test, t_0 is a root of the non-linear equation

$$\frac{1}{n}\sum_{j=1}^{m}\frac{f_j^2}{p_j(t)} - n = \chi_{1-\alpha}^2,$$

where the left-hand side is the chosen testing criterion and $\chi_{1-\alpha}^2$ is a $(1 - \alpha)$-quantile of the chi-squared distribution with $m - 1$ degrees of freedom. The problem here is to find a value of t that still does not lead to rejecting the hypothesis of $p(t)$ being a good fit at a significance level of α.

All the gradient estimates $p(t_0)$ for $\forall t \in [0; t_0]$ meet the chosen testing criterion at a significance level of at least α and the estimate $p(t_0)$ being the "worst" of all estimates, can be referred to as a ***pessimistic gradient estimate***.

In the event that we would also like to obtain an interval estimate it is possible to utilize the bootstrap method. From random sample $(x_1, \ldots, x_n)$ a new random sample $(x_1^*, \ldots, x_n^*)$ with repetition (with return) was created. We selected new bootstrap samples from the statistic sample observed values x_i by random draw with replacement.

The random sample so obtained is denoted as a bootstrap sample. It is possible to create the bootstrap interval estimate in the following way (Efron and Tibshirani, 1994).

Let $\overline{x}$ be the mean and s^2 be the variance of original sample $(x_1, \ldots, x_n)$, $\overline{x}_{b,i}$ is the mean and $s_{b,i}^2$ is the variance of the i-bootstrap sample, $i = 1, \ldots, B$, then

1. The bootstrap interval estimate of the mean $E(X)$ with a confidence level $1 - \alpha$ is

$$\left\langle \overline{x} - t_{b,1-\frac{\alpha}{2}}\frac{s}{\sqrt{n-1}},\ \overline{x} - t_{b,\frac{\alpha}{2}}\frac{s}{\sqrt{n-1}} \right\rangle,$$

where $t_{b,P}$ is the *P*-quantile of the sample $(t_{b,1}, \ldots, t_{b,B})$ and

$$t_{b,i} = \frac{\overline{x}_{b,i} - \overline{x}}{s_{b,i}}\sqrt{n-1},\ i = 1, \ldots, B.$$

2. The bootstrap interval estimate of the variance $D(X)$ with a confidence level $1 - \alpha$ is

$$\left\langle \frac{ns^2}{\chi_{b,1-\frac{\alpha}{2}}^2},\ \frac{ns^2}{\chi_{b,\frac{\alpha}{2}}^2} \right\rangle,$$

where $\chi_{b,P}^2$ is the *P*-quantile of the sample $(\chi_{b,1}^2, \ldots, \chi_{b,B}^2)$ and

$$\chi_{b,i}^2 = \frac{ns_{b,i}^2}{s^2},\ i = 1, \ldots, B.$$

3. The bootstrap interval estimate of the standard deviation $\sigma(X)$ with a confidence level $1 - \alpha$ is calculated as the square root of the variance $D(X)$.

RESULTS AND DISCUSSION

The Pareto diagram is an important tool for managerial decision-making. The diagram helps to separate important factors from less important ones. Applying a gradient estimate reduces the difference between the factors. If the differences between the factors is sufficient, the result of the Pareto analysis remains the same as that for empirical frequencies. Conversely, if the difference is not too great, the Pareto analysis will point to there being more important factors.

Application of a Pessimistic Gradient Estimate

In a manufacturing plant, an inconsistency in the output products was observed. In order to improve the quality of production, the types and numbers of defects of individual products were noted over one month. To assess the types of defects which affect the product consistency the most, the Pareto analysis was performed. It was based on both the output quality inspection data and data estimated

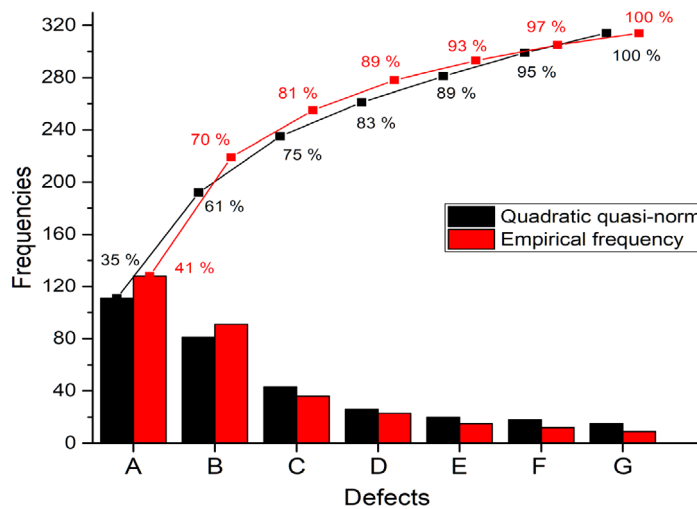I: *Empirical and estimated frequencies for $t_0 = 0.0565$*

Defect	A	B	C	D	E	F	G
Frequencies of defects	128	91	36	23	15	12	9
Cumulative frequencies	41%	70%	81%	89%	93%	97%	100%
Estimates of rel. frequencies	0.3535	0.2580	0.1369	0.0828	0.0637	0.0573	0.0478
Estimated frequencies	111	81	43	26	20	18	15
Cumulative frequencies	35%	61%	75%	83%	89%	95%	100%

using Quadratic quasi-norm. The results are shown in a Tab. I and in the Fig. 1.

If data estimated using a quadratic quasi-norm are evaluated using a Pareto analysis, the company should eliminate defects A-D to improve production. The estimations given are point estimations. The bootstrap method can be to obtain interval estimations. The procedure is as follows. A thousand bootstrap samples are generated and the estimate is calculated from each resample using a quadratic quasi-norm. Each estimate is characterized by a parameter t. The value of t ranges between 0 and 1.

We determine from these set of parameters value t_0 (pessimistic estimate). The probabilities p_i were calculated from value $t_0 = 0.565$. Their numerical characteristics are shown in Tab. II.

Subsequently, the bootstrap was performed again, from which the interval estimate of the mean, variance and standard deviation for the probabilities p_i was obtained. These are estimates of individual probabilities with the confidence level 0.95. The simultaneous estimate is unknown. The results are shown in Tab. III.

1: *Empirical and estimated frequencies*

II: *Numerical characteristics of bootstrap samples of probabilities*

Def.	Mean	Median	Min.	Max.	Lower Q.	Upper Q.	Variance	Std. Dev.
A	0.371504	0.372358	0.291666	0.466385	0.352341	0.391046	0.000761	0.027584
B	0.274932	0.274172	0.199755	0.361052	0.258063	0.290117	0.000559	0.023639
C	0.109539	0.109735	0.059762	0.174739	0.098138	0.118811	0.000269	0.016407
D	0.082311	0.082949	0.044394	0.117565	0.074330	0.091145	0.000155	0.012431
E	0.072207	0.071751	0.023652	0.109717	0.062991	0.080943	0.000167	0.012909
F	0.048943	0.048402	0.018613	0.086108	0.040560	0.056575	0.000129	0.011343
G	0.040566	0.040082	0.013143	0.072303	0.033348	0.046859	0.000092	0.009577

III: *Bootstrap 95% confidence intervals of probabilities*

BOOTSTRAP 95% CONFIDENCE INTERVALS					
Defect	MEAN		VARIANCE		STD. DEV.	
A	0.369848	0.373165	0.000721	0.000856	0.026860	0.029256
B	0.272943	0.276847	0.000120	0.000151	0.010934	0.012272
C	0.108476	0.110618	0.000256	0.000315	0.015994	0.017759
D	0.081613	0.083183	0.000145	0.196403	0.012027	0.443174
E	0.071095	0.073498	0.000158	0.000187	0.012560	0.013676
F	0.048352	0.050059	0.000122	0.000144	0.011028	0.011987
G	0.039441	0.041377	0.000086	0.000101	0.009270	0.010061

CONCLUSION

There are many practical situations where it is necessary to estimate the probability distribution of observed categorical variables. They occur in industry and in empirical research, in life-time testing, fault diagnosis and in social surveys. The article deals with the current issue of obtaining estimates of the parameters of the distribution of categorical random variables. The initial apparatus is the f-divergence of discrete probability models. The concept of a quasinorm is derived from them, i.e. the divergence of a specific distribution estimate from a discrete uniform distribution.

Due to the variability of the parameter t choice from the interval $[0; t_0]$, they are sufficiently flexible and allow an approximation to the observed empirical distribution. It is possible to choose different f-quasinorms and obtain other variants of pessimistic estimates (Karpíšek *et al.*, 2008). However, this generally leads to a nonlinear system of differential equations and requires its numerical solution. This quadratic f-quasinorm is the only one that corresponds to a linear system of differential equations with constant coefficients and whose solution can be obtained without numerical methods.

By solving the aforementioned SODR1 we get estimates directly of the probabilities being sought. Since f-quasinorms are symmetric functions, it is possible to replace condition $p_m(t) = 1 - \sum_{i=1}^{m-1} p_i(t)$ with condition $p_k(t) = 1 - \sum_{i=1, i\neq k}^{m} p_i(t)$. The numbering of categorical variables can in fact be arbitrary. Another benefit for this issue is the use of bootstrapping, especially in the construction of interval estimates. This method is universal and is applicable to other quasinorms as well as to other types of discrete probability distribution estimates. In the case of other quasinorms, the calculation is more difficult, because the nonlinear system of differential equations must be solved again for each bootstrap selection. The generation of bootstrap files and the necessary calculations were performed in software R.

Gradient estimates based on f-quasinorms have a non-traditional and specific character, but a number of other tasks have demonstrated their practical applicability, e.g. in category analysis, with applications also outside sociological research (Karpíšek and Neradová, 2014; Lacinová *et al.*, 2011).

REFERENCES

EFRON, B. and TIBSHIRANI, R. J. 1994. *An Introduction to the Bootstrap*. Boca Raton: Chapman & Hall/CRC.

KARPÍŠEK, Z., JURÁK, P. and ŠÁCHA, J. 2007. Divergences for Discrete Probability Distribution Estimations. In: *Summer School DATASTAT '06*. Proceedings. Brno: Masaryk University, pp. 287–292.

KARPÍŠEK, Z., NERADOVÁ (LACINOVÁ), V. and ZAMPACHOVÁ, E. 2008. A Contribution to the Estimation of Discrete Probability Distribution. In: *MENDEL 2008. 14th International Conference on Soft Computing* (pp. 287–292). Brno.

KARPÍŠEK, Z. and NERADOVÁ (LACINOVÁ), V. 2014. Gradient and Line Estimates Employed in Surveyes. In *MENDEL 2014. 20th International Conference on Soft Computing.* Brno: FSI, VUT Brno, pp. 407–414.

LACINOVÁ, V., KARPÍŠEK, Z. and SADOVSKÝ, Z. 2011. Pessimistic estimates of the probability distribution of a categorical quantity [in Czech: Pesimistické odhady rozdělení pravděpodobnosti kategoriální veličiny]. *Informační bulletin České statistické společnosti*, 22(2): 138–145.

VAJDA, I. 2004. *Theory of Information* [in Czech: *Teorie informace*] Praha: ČVUT.

Contact information

Jakub Šácha: jakub.sacha@mendelu.cz