УЧЕНЫЕ ЗАПИСКИ КАЗАНСКОГО УНИВЕРСИТЕТА.
СЕРИЯ ФИЗИКО-МАТЕМАТИЧЕСКИЕ НАУКИ

2018, Т. 160, кн. 2
С. 410–418

ISSN 2541-7746 (Print)
ISSN 2500-2198 (Online)

UDK 519.226.3

d-POSTERIOR APPROACH IN REGRESSION

A.A. Zaikin

Kazan Federal University, Kazan, 420008 Russia

Abstract

In this paper, we have used the d-posterior approach in regression. Regression predictions are a sequence of similarly made decisions. Thus, d-risk can be helpful to estimate the quality of such decisions. We have introduced a method to apply the d-posterior approach in regression models. This method is based on posterior predictive distribution of the dependent variable with the given novel input of predictors. In order to make d-risk of the prediction rule meaningful, we have also considered adding probability distribution of the novel input to the model.

The method has been applied to simple regression models. Firstly, linear regression with Gaussian white noise has been considered. For the quadratic loss function, estimates with uniformly minimal d-risks have been constructed. It appears that the parameter estimate in this model is equal to the Bayesian estimate, but the prediction rule is slightly different. Secondly, regression for the binary dependent variable has been investigated. In this case, the d-posterior approach is used for the logit regression model. As for the 0-1 loss function, the estimate with uniformly minimal d-risk does not exist, we suggested a classification rule, which minimizes the maximum of two d-risks. The resulting decision rules for both models are compared to the usual Bayesian decisions and the decisions based on the maximum likelihood principle.

Keywords: Bayesian inference, regression, d-risk

Introduction

When solving the regression problem, the first step is always "training" the model using a finite sample. Only then are the estimates of the model parameters used to predict the dependent variable value for every new set of predictors. The prediction itself is a process of making a variety of similar decisions, which gives a reason to apply the d-posterior approach [1, 2] to control potential risks of such decisions. Since d-risk can be interpreted as expected loss for a particular prediction value, the d-posterior approach is a natural alternative to the maximum likelihood and Bayesian principles in solving this problem.

The d-posterior approach has not been tried for regression problems yet. That is to say, there is a regression technique [3], which controls the false discovery rate (FDR), the decision quality related to d-risk. However, the statistical model used in [3] is not the same as in the classical regression problem statement. Here, we discuss two classical regression models: linear regression with Gaussian white noise and quadratic loss function, and logit linear regression with zero-one loss function.

1. Regression

All vectors will be treated as column vectors in this paper. We will also denote transposing with a superscript T, so that X^TY is a scalar product of two real m-vectors for $X \in \mathbb{R}^m$ and $Y \in \mathbb{R}^m$.

Let us fix positive integers m and n, such that $n \geq m$, and suppose that we have a full rank real matrix

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & \dots & x_{1,m} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \dots & x_{n,m} \end{bmatrix}.$$

This matrix defines predicates for regression. Let X_i be i-th row of the matrix $\mathbf{X}$.

We assume that for every vector $X \in \mathbb{R}^m$ there exists some density $p(y|X,\theta)$ with respect to some measure, where θ is an unknown parameter from the parameter space Θ. In this paper, we consider only the case of linear regression, where $\Theta = \mathbb{R}^m$ and $p(y|X,\theta) = f(y|X^T\theta)$ for some density f. Hereafter, any kind of probability density would be denoted as p. It will be clear from the context which density is meant.

The d-posterior approach is a subsection of the Bayesian statistics. Therefore, we need to define the prior distribution of the parameter θ. We assume that θ is a sample value of the continuous random variable ϑ with the known density $p(\theta)$.

Let us suppose that we deal with independent observations $Y = (y_1, \dots, y_n)^T$, which follow the distribution defined by $p(y|X_i,\theta)$. The likelihood function is $p(Y|\mathbf{X},\theta) = \prod_{i=1}^{n} p(y_i|X_i,\theta)$. The posterior density is

$$p(\theta|Y,\mathbf{X}) = p(Y|\mathbf{X},\theta)p(\theta) \Big/ \int\limits_{\Theta} p(Y|\mathbf{X},\theta)p(\theta)\,d\theta. \tag{1}$$

The first problem with regression is to find some estimate $\widehat{\theta}$ of θ, which can be used later to estimate the density $p(y_*|X_*,\widehat{\theta})$ of the predicted variable y_* for a novel predictor input X_*. The usual maximum likelihood estimate (MLE) maximizes the likelihood $p(Y|\mathbf{X},\theta)$ with respect to θ. Another option is to use the maximum a posteriori (MAP) estimate, which maximizes the posterior density instead (the same as maximizing $p(Y|\mathbf{X},\theta)p(\theta)$). Both options are very popular (MAP estimates are used in numerous regularization techniques, such as ridge regression, LASSO, etc.).

The less popular option is the Bayesian estimate. Firstly, one needs to specify the loss function $L(\theta_1,\theta_2)$. The prior risk function for estimate $\widehat{\theta}$ is then $\mathcal{R}_{\widehat{\theta}} = \mathbf{E}L(\vartheta,\widehat{\theta})$. The Bayesian estimate minimizes the risk with respect to $\widehat{\theta}$. The corresponding estimation could be achieved more easily by minimizing the posterior risk

$$R(d|Y,\mathbf{X}) = \mathbf{E}\left[L(\vartheta,d) \;\middle|\; Y,\mathbf{X}\right] = \int\limits_{\Theta} L(\theta,d)p(\theta|Y,\mathbf{X})\,d\theta$$

with respect to d. In the case of the quadratic loss function, the Bayesian estimate is a posterior mean. The Bayesian estimates are not widely used, because they usually require intensive computation.

The d-risk of the estimate $\widehat{\theta}$ is

$$\mathcal{R}_{\widehat{\theta}}(d) = \mathbf{E}\left[L(\vartheta,\widehat{\theta})|\widehat{\theta} = d\right].$$

Since d-risk is a function, there are different possible ways to define the estimate that "minimizes" its d-risk. For some statistical models, this minimizing is trivial: there exists such an estimate $\widehat{\theta}^*$ that

$$\mathcal{R}_{\widehat{\theta}^*}(d) \leq \mathcal{R}_{\widehat{\theta}}(d), \tag{2}$$

for any d and any estimate $\widehat{\theta}$. The estimate $\widehat{\theta}^*$, which satisfies (2), is called an estimate with uniformly minimal d-risk (we will designate it as U-estimate). There is a way to

find this estimate [1]: one needs to minimize $R(d|Y,\mathbf{X})$ with respect to its random arguments (in the case of the above-given regression statement, it is Y). If for every Y exists at least one d, for which $R(d|Y,\mathbf{X})$ is minimized, then this d (or any, if there are multiple solutions) is a U-estimate. As it was said earlier, none of such estimates have been used in the literature.

However, estimating the parameters is not best suited for the d-posterior approach. The thing is, the d-risk can be interpreted as average loss for a particular decision value among a succession of experiments. Actually, the parameter θ is estimated only once. Therefore, U-estimates can be viewed only as somewhat regularized MLEs, just like the Bayesian or MAP estimates. On the other hand, prediction for a single training sample can be made infinitely many times. This gives the opportunity to use the d-risk for assessing the quality of custom prediction rules. However, in order to make the most sense of the definition of d-risk, we need to make the predictors random. Indeed, in the case of constant predictor vector X_*, the d-risk is "average loss for particular decision value among a succession of experiments, *with predictors equal to* X_*". If X_* is a random vector, then d-risk is "average loss for particular decision value among a succession of experiments". Hence, we will present various possibilities for X_* distribution. Note that we formally do not need to specify the distribution of predictors of the matrix $\mathbf{X}$.

For a new set of predictors X_*, the predictive posterior distribution of the dependent variable y_* is

$$p(y_*|X_*,Y,\mathbf{X}) = \int\limits_{\Theta} p(y_*|\theta,X_*)p(\theta|\mathbf{X},Y)\,d\theta. \tag{3}$$

The right side of (3) is obtained using the fact that distribution of y_* provided that X_* and θ are not dependent on the training sample Y and $\mathbf{X}$. The same can be said about the distribution of Y, which does not depend on X_* given $\mathbf{X}$ and θ. The posterior predictive of y_* can be used in the same manner as the usual posterior distribution in terms of the Bayesian rules and rules which minimize the d-risk. In this case, y_* is the "parameter", for which we can specify the loss function and posterior risk.

2. Linear regression with Gaussian noise

This section focuses on studying of the following model:

$$y_i = X_i^T\theta + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0,\sigma^2), \quad i = 1,\dots,n, \tag{4}$$

for a mutually independent ε_i. We assume the variance σ^2 of the white noise to be known. We will discuss this matter later.

The likelihood function of this model is

$$p(Y|\mathbf{X},\theta) = \varphi\left(Y|\mathbf{X}\theta,\sigma^2 I_n\right),$$

where $\varphi(x|\mu,\Lambda)$ is the density function of multivariate normal distribution with the mean vector μ and covariance matrix Λ, and I_d is the $d\times d$ identity matrix. The corresponding cumulative distribution function will be denoted as $\Phi(x|\mu,\Lambda)$. For the sake of simplicity, the PDF and CDF for the standard univariate (with mean zero and variance equal to one) normal distribution will be denoted as $\varphi(x)$ and $\Phi(x)$, respectively.

For this model, the MLE of θ is well-known:

$$\widehat{\theta}_{ml} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Y.$$

If we try to apply the maximum likelihood principle to prediction, we can obtain the usual predicting scheme, which involves MLE of θ. Indeed, in order to maximize

$p(y^*, Y|\theta, \mathbf{X}, X_*)$ with respect to unknown parameters, we need to maximize it with respect to θ and y_*. Since the mode of normal distribution is its mean, the MLE of y_* is $X_*^T\widehat{\theta}_{ml}$. Note that here we do not need to estimate σ^2 for prediction.

For the Bayesian analysis, we need to specify the prior:

$$p(\theta) = \varphi(\theta|0, \tau^2 I_m).$$

The parameter τ is known.

The posterior density of θ is

$$p(\theta|Y, \mathbf{X}) = \varphi\left(\theta\Big|\frac{S\mathbf{X}^TY}{\sigma}, S\right), \quad S = \left(\frac{I_m}{\tau} + \frac{\mathbf{X}^T\mathbf{X}}{\sigma}\right)^{-1}.$$

This expression can be derived from the convolution theorem. The Bayesian estimate of θ for the quadratic loss function $L(\theta, d) = \|\theta - d\|_2^2$ is the posterior mean:

$$\widehat{\theta}_B = \frac{S\mathbf{X}^TY}{\sigma}.$$

Note that this expression depends on σ and τ.

In the same Bayesian setting, the posterior predictive for a new set of the predictor variables X_* is

$$p(y_*|X_*, Y, \mathbf{X}) = \varphi\left(y_*\Big|\frac{X_*^TS\mathbf{X}^TY}{\sigma}, X_*^TSX_* + \sigma^2\right). \tag{5}$$

This can be derived from the direct representation of the model (4), because the linear transformation of normal distribution is normal distribution as well. If the posterior predictive is treated as posterior distribution, then for the quadratic loss function $L(y_*, d) = (y_* - d)^2$ the Bayesian estimate for y_* is $X_*\widehat{\theta}_B$.

Now, let us consider the d-posterior approach to estimate θ. For the model studied in this section, we can find the U-estimate. Indeed, the posterior risk can be expressed as

$$\begin{aligned} R(d|Y, \mathbf{X}) &= \mathbf{E}\left[\|\vartheta - d\|_2^2 \;\middle|\; Y, \mathbf{X}\right] = \\ &= \mathbf{E}\left[\|\vartheta - \widehat{\theta}_B\|_2^2 \;\middle|\; Y, \mathbf{X}\right] + \|\widehat{\theta}_B - d\|_2^2 = \operatorname{trace}(S) + \|\widehat{\theta}_B - d\|_2^2. \end{aligned}$$

We used the fact that $\mathbf{E}\left[\vartheta \;\middle|\; Y, \mathbf{X}\right] = \widehat{\theta}_B$. Since the only term which depends on Y is the term with $\widehat{\theta}_B$, the minimizing with respect to Y is straightforward, and the U-estimate of θ is equal to the Bayesian estimate $\widehat{\theta}_B$.

Now, we want to find the U-estimate for y_*. As it was said in the previous section, in order to make the use of d-risk meaningful, we need to consider X_* random. Surprisingly, we do not need to specify it unless we assume that it does not depend on θ. Then, the posterior predictive distribution does not depend on the distribution of X_*, and it is given by formula (5). On a side note, if we consider $\mathbf{X}$ to be random with distribution which does not depend on θ, then the posterior predictive also does not change.

In order to find the U-estimate of y_* for the quadratic loss function, we need to minimize

$$R(d|X_*, Y, \mathbf{X}) = \mathbf{E}\left[(y_* - d)^2 \;\middle|\; X_*, Y, \mathbf{X}\right]$$

with respect to X_*, Y. Unfortunately, it seems that there is no invertible solution of that optimization problem. However, if we consider only Y to be random, the U-estimate of y_* is equal to the Bayesian estimate. Considering only X_* to be random

gives a completely different result. Indeed, the posterior risk of the decision d can be expressed as

$$R(d|X_*, Y, \mathbf{X}) = \mathbf{E}\left[(y_* - X_*^T\widehat{\theta}_B)^2 \mid X_*, Y, \mathbf{X}\right] + (X_*^T\widehat{\theta}_B - d)^2 = \\ = X_*^T S X_* + \sigma^2 + (X_*^T\widehat{\theta}_B - d)^2.$$

Differentiating with respect to X_* yield

$$R'(d|X_*, Y, \mathbf{X}) = 2SX_* + 2\widehat{\theta}_B\left(X_*^T\widehat{\theta}_B - d\right).$$

Now, we need to solve the equation $R'(d|X_*, Y, \mathbf{X}) = 0$. By performing some transformations, we can get a solution for d:

$$2SX_* + 2\widehat{\theta}_B\left(X_*^T\widehat{\theta}_B - d\right) = 0,$$

$$\widehat{\theta}_B^T SX_* + \widehat{\theta}_B^T\widehat{\theta}_B\left(X_*^T\widehat{\theta}_B - d\right) = 0,$$

$$d = \frac{\widehat{\theta}_B^T SX_*}{\widehat{\theta}_B^T\widehat{\theta}_B} + X_*^T\widehat{\theta}_B = X_*^T\frac{S\widehat{\theta}_B}{\widehat{\theta}_B^T\widehat{\theta}_B} + X_*^T\widehat{\theta}_B.$$

The right side of the last expression is obtained by transposing the first term on the left side. This yields the linear dependence on X_* of the U-estimate for y_*: $y_* = X_*^T\widehat{\theta}_U$, where

$$\widehat{\theta}_U = \widehat{\theta}_B + \frac{S\widehat{\theta}_B}{\widehat{\theta}_B^T\widehat{\theta}_B}.$$

The fixed values of Y do not significantly affect interpretation. For the fixed Y, the d-risk is an "average loss for particular decision value among a succession of experiments, *given the training set* Y". One can argue that this interpretation is even more natural and useful than the interpretation of d-risk for a random Y.

The problematic part here is the fact that hyperparameters σ and τ need to be known. In case of unknown hyperparameters, the ML-II procedure (maximum likelihood estimates) is popular. Another approach is to specify the distributions of σ and τ, and use the marginal likelihood in all derivations. This approach is useful because marginal likelihoods obtained that way are usually not very sensitive to changes of second-level prior distributions parameters. Unfortunately, all these techniques do not yield results which can be expressed with explicit formulas. It is also very difficult to apprehend their impact on the posterior distribution and U-estimates.

3. Logit linear regression

In this section, we assume that dependent variables take only two values (0 and 1) and follow the logit linear regression model. The density of a single observation is

$$\mathbf{P}(y_i = 1|\theta, X_i) = \sigma\left(X_i^T\theta\right),$$

where sigmoid function is given by

$$\sigma(x) = \frac{1}{1 + e^{-x}}.$$

The likelihood function in this case is given by

$$p(Y|\theta, \mathbf{X}) = \prod_{i=1}^{n}\sigma\left((2y_i - 1)X_i^T\theta\right).$$

The prior of θ is assumed to be Gaussian: $p(\theta) = \varphi(\theta|0, \tau^2)$. Here, we also consider X_* to be random, and its distribution does not depend on θ. One such possibility is to consider X_* Gaussian with the mean vector 0 and covariance matrix Λ.

Unfortunately, integrals and optimization problems are intractable. Binary linear regression is usually fitted by the numerical methods, regardless of the method of estimating (MLE and MAP estimates or Bayesian estimate).

The posterior density of θ is given by (1). The posterior predictive of a new set of predictors X_* is given by (3). Now, let us suppose that the loss function is given by

$$L(y_*, d) = \begin{cases} 1, & y_* = d, \\ 0, & y_* \neq d. \end{cases}$$

The Bayesian prediction rule for L predicts the value k that has the maximum predictive posterior probability of $\mathbf{P}(y_* = k|X_*, Y, \mathbf{X})$.

The next step is to predict the value of y_* given the new input X_*. For this purpose, we define the classification rule $\phi = \phi(X_*, Y, \mathbf{X})$, so the values of ϕ correspond to predicted values of y_*. D-risks of the classification rule ϕ for the 0-1 loss function are given by

$$\mathcal{R}_\phi(0) = \mathbf{P}(y_* = 1|\phi = 0), \quad \mathcal{R}_\phi(1) = \mathbf{P}(y_* = 0|\phi = 1).$$

These formulas can be expressed as

$$\mathcal{R}_\phi(k) = \int\limits_{\Theta} \int\limits_{\mathbb{R}^m} \mathbf{P}(y_* = 1 - k|\theta, X_*) p(\theta, X_*|\phi = k)\, dX_*\, d\theta,$$

where

$$p(\theta, X_*|\phi = k) = \mathbf{P}(\phi = k|\theta, X_*) p(\theta) p(X_*) \Big/ \int\limits_{\Theta} \int\limits_{\mathbb{R}^m} \mathbf{P}(\phi = k|\theta, X_*) p(\theta) p(X_*)\, dX_*\, d\theta.$$

This expression for d-risk is convenient, because we usually know the distribution $\mathbf{P}(\phi = k|\theta, X_*)$.

In the case of finite decision spaces, U-estimates do not exist in most cases [1], so we need to use different definition of the optimal decision rule. The binary classification can be perceived as a problem of comparing two hypotheses. There exists [4] such a rule ϕ^* that for the given $0 < \beta_0 < 1$ $\mathcal{R}_{\phi^*}(0) \leq \beta_0$ and $\mathcal{R}_{\phi^*}(1)$ is minimal among all the rules ϕ that satisfy $\mathcal{R}_\phi(0) \leq \beta_0$. The most important thing here is that such a rule (in setting of the prediction problem of the current section) has the following form:

$$\phi^* = \begin{cases} 1, & T > C, \\ 0, & T \leq C \end{cases},$$

for some constant C and $T = \mathbf{P}(y_* = 1|X_*, Y, \mathbf{X})$. Note that for the Bayesian prediction rule $C = 0.5$.

Of course, one can define β_0 and numerically find C, for which $\mathcal{R}_{\phi^*}(0) \leq \beta_0$. This is the case when one decision is more important than other. However, usually researchers do not distinguish different values of dependent variables, and the most natural way to find such C is that $\mathcal{R}_{\phi^*}(0) = \mathcal{R}_{\phi^*}(1)$. It is possible for d-risks which are small enough, or, conversely, large n. The latter is due to the fact that d-risks tend to 0 as $n \to \infty$, see [5].

Note that we can calculate d-risks of ϕ for the fixed X_* using

$$\mathcal{R}_\phi(k|X_*) = \mathbf{P}(y_* = 1 - k|\phi = k, X_*), \quad k = 0, 1. \tag{6}$$

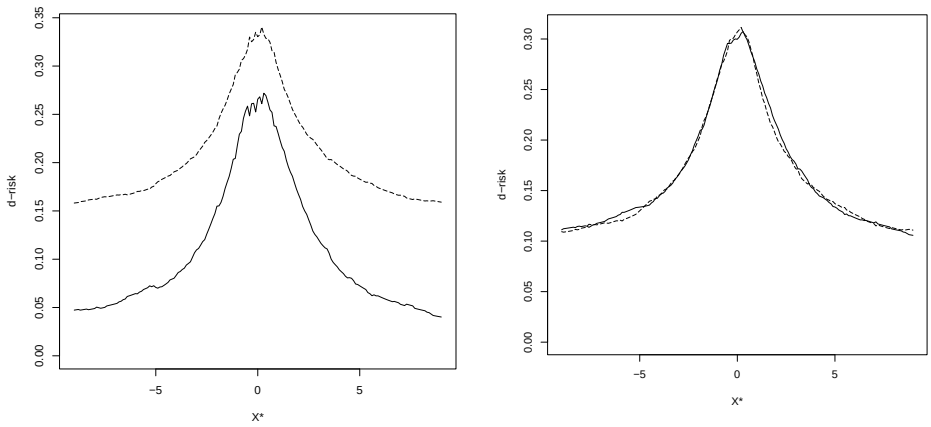

Fig. 1. Sample d-risks (6) of ϕ^* for $C = 0.4$ and $C = 0.5$. Solid line is $\mathcal{R}_{\phi^*}(0)$ and dashed line is $\mathcal{R}_{\phi^*}(1)$

In order to get $\mathcal{R}_\phi(k)$, one needs to calculate expectation of the last expression with respect to X_*.

The expression (6) is very interesting as it shows which values of X_* are the most risky in taking decisions. In order to see that, we set up a numerical example. We consider a simple regression model of the form

$$\mathbf{P}(y_i = 1|\theta, X_i) = \sigma\left(\theta_0 + x_i\theta_1\right),$$

where x_i is a scalar value. Let $(\theta_0, \theta_1)^T \sim \mathcal{N}(0, 4I_2)$, $x_i, x_* \sim \mathcal{N}(0, 9)$, $i = 1, \ldots, n$, $n = 25$. In the numerical Monte-Carlo experiment, we calculated frequencies of errors and calculated conditional d-risks (6). The results for ϕ^* with $C = 0.4$ and $C = 0.5$ are shown in Fig. 1. Due to the symmetry of distributions of all variables, it is expected that d-risk has also a symmetrical form. The values of x_* with the largest d-risk are those closest to zero. It is also expected, because then y_* takes values 0 and 1 with the expected probability of 0.5.

Conclusions

We studied the ways to construct the prediction rules, which minimize the d-risk. For the linear regression with Gaussian noise, we constructed the U-estimate for the regression parameter θ and U-estimate for prediction of y_* of a new observation X_* for random X_* and fixed Y. For the linear logit regression, we suggested the prediction rule, which minimizes the maximum of two d-risks.

Acknowledgements. This work was funded by the subsidy allocated to Kazan Federal University for the state assignment in the sphere of scientific activities (project no. 1.7629.2017/8.9) (for Gaussian regression). The study was also supported by the Russian Foundation for Basic Research and the Republic Of Tatarstan according to the research project no. 17-41-160620 (for logit regression).

The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

References

1. Volodin I.N., Simushkin S.V. On d-posteriori approach to the problem of statistical inference. *Proc. 3rd Int. Vilnius Conf. on Probability Theory and Mathematical Statistics*, 1981, vol. 1, pp. 100–101.
2. Volodin I.N., Simushkin S.V. Statistical inference with minimal d-risk. *J. Sov. Math.*, 1988, vol. 42, no. 1, pp. 1464–1472. doi: 10.1007/BF01098858.
3. Scott J.G., Kelly R.C., Smith M.A., Zhou P., Kass R.E. False discovery rate regression: An application to neural synchrony detection in primary visual cortex. *J. Am. Stat. Assoc.*, 2015, vol. 110, no. 510, pp. 459-471. doi: 10.1080/01621459.2014.990973.
4. Simushkin S.V. Optimal d-guarantee procedures for distinguishing two hypothesis. *VINITI Acad. Sci. USSR*, 1981, no. 55, pp. 47–81. (In Russian)
5. Volodin I.N., Novikov A.A. Asymptotics of the necessary sample size in testing parametric hypotheses: d-posterior approach. *Math. Methods Stat.*, 1998, vol. 7, no. 1, pp. 111–121.

Received
October 12, 2017

Zaikin Artyom Alexandrovich, Assistant of the Department of Mathematical Statistics
Kazan Federal University
ul. Kremlevskaya, 18, Kazan, 420008 Russia
E-mail: *kaskrin@gmail.com*

УДК 519.226.3

d-Апостериорный подход в регрессии

А.А. Заикин

Казанский (Приволжский) федеральный университет, г. Казань, 420008, Россия

Аннотация

В статье представлена попытка применить d-апостериорный подход в регрессии. Так как регрессионные прогнозы являются по сути последовательностью схожих решений, это даёт возможность использования d-риска как меры качества прогнозирования. В работе изучаются различные подходы к применению d-апостериорного подхода для прогноза в регрессионных моделях. Предлагается подход, основанный на апостериорном прогностическом распределении переменной-регрессора в зависимости от значений переменных-предикторов. Для того чтобы интерпретация d-риска правила прогноза имела смысл, предлагается добавить в вероятностную модель распределение предикторов.

Эта методика была применена на двух простых регрессионных моделях. Сначала изучается линейная регрессия с гауссовским белым шумом. Для этой модели и для квадратической функции потерь были построены оценки с равномерно минимальным d-риском. Оказалось, что оценка параметра совпадает с байесовской оценкой, а прогноз несколько отличается. Далее рассматривается логистическая регрессия для бинарной зависимой переменной. Для функции потерь 1–0 не существует правила прогоноза, равномерно минимизирующего d-риск, поэтому предлагается правило, которое минимизирует максимум двух d-рисков. Полученные для обеих моделей правила сравниваются с известными решающими функциями, построенными согласно Байесовскому принципу и принципу максимального правдоподобия.

Ключевые слова: байесовская статистика, регрессия, d-риск

Поступила в редакцию
12.10.17

Заикин Артём Александрович, ассистент кафедры математической статистики
Казанский (Приволжский) федеральный университет
ул. Кремлевская, д. 18, г. Казань, 420008, Россия
E-mail: *kaskrin@gmail.com*

⟨ ***For citation*:** Zaikin A.A. d-Posterior approach in regression. *Uchenye Zapiski Kazanskogo Universiteta. Seriya Fiziko-Matematicheskie Nauki*, 2018, vol. 160, no. 2, pp. 410–418. ⟩

⟨ ***Для цитирования*:** *Zaikin A.A.* d-Posterior approach in regression // Учен. зап. Казан. ун-та. Сер. Физ.-матем. науки. – 2018. – Т. 160, кн. 2. – С. 410–418. ⟩