ON AN APPLICATION OF BOOTSTRAP TO THE ANALYSIS OF ANIMAL GROWTH

POUŽITÍ METODY BOOTSTRAPU K ANALÝZE RŮSTOVÝCH DAT

Helena Nešetřilová

Summary:

Statistical analysis of animal breeding data often requires modelling of animal growth and body composition. Models which are used for this analysis are mainly nonlinear regression models. Classical methods of estimation of nonlinear regression model parameters do not allow to find the properties of the nonlinear model parametr estimates. As a robust alternative, the bootstrap method can be used to find the precision of the estimates and their correlations.

Anotace:

Statistická analýza dat v oblasti živočišné výroby je často založena na konstrukci růstových modelů a modelů charakterizujících vzájemné vztahy mezi složením jatečného těla zvířat. Modely, které jsou k tomuto účelu používány, jsou obvykle nelineární regresní modely. Klasické metody odhadu parametrů těchto modelů neumožňují zjistit dostatečně přesně skutečné vlastnosti získaných odhadů. Přesnost odhadu parametrů těchto modelů je možno ověřit pomocí metody bootstrapu. Tato metoda také umožňuje odhadnout korelace mezi parametry modelu.

Key words:

growth curves, Gompertz function, bootstrap, cattle growth

Klíčová slova:

růstové křivky, Gompertzova funkce, bootstrap, růst skotu

Animal breeders are often interested in statistical analysis of data which are observed during growth process of an individual animal or a group of animals (e.g. breed). This task usually leads to a construction of a suitable growth model. On general, growth models are nonlinear regression models.

Suppose that such a model (nonlinear in its parameters) is

$$y_i = f(x_i, \boldsymbol{\theta}) + \varepsilon_i, \; i = 1, \dots, n, \qquad (1)$$

where $\boldsymbol{\theta}$ is a vector of parameters, ε_i are independent identically distributed random errors, for which expected value $E(\varepsilon_i) = 0$ and variance $\sigma^2 > 0$, and the first and second derivates of function $f(x_i, \boldsymbol{\theta})$ by all components of the vector $\boldsymbol{\theta}$ and x_i are continuous. Usually, we also assume that random errors ε_i have the normal distribution.

Estimation of the parameters of the model (1) is usually based on the least square criterion which leads to a systém of nonlinear equations for the parameter estimates which has to be solved iterativelly. As far as the properties of the estimates are concerned, only asymptotic properties (e.g. for infinite sample size) are known. Under the assumptions made above (including normality), the least squares (LS) estimators are also maximum likelihood

(ML) estimators. Asymptotic theory tells (see e.g. Ratkowsky, 1983, p.6) that with increasing sample size the ML estimator becomes more and more unbiased, more and more normally distributed and approaches a minimum possible variance (called the minimum variance bound). For finite samples, it is not clear how large must the sample size be for these asymptotic properties to be closely approximated. Thus, for small finite samples, the LS estimates of the model (1) parameters are not in general normally distributed, nor unbiased nor minimum variance. This makes it also difficult to evaluate precision of estimates (their standard error) and to set bounds for their confidence intervals.

Thus, the alternative methods to help with the above problems are welcome. An effective method of this kind is bootstrap. The bootstrap (proposed by J. Simon in 1969 independently by B. Efron) is a general technique based on resampling (with replacement) the observed sample data. It has many variants, but its basic idea is to approximate the real sampling distribution of statistics used to construct confidence intervals or to test hypotheses (see e.g. Wilcox, 2001). The bootstrap technique can give highly accurate results in situations where more traditional methods perform poorly.

In this paper, the bootstrap method is used to supplement the traditional statistical analysis of a growth model. The aim of the statistical analysis was to construct the Gompertz growth model

$$y(t) = \alpha \exp(-\exp(\beta - \gamma t)) \qquad (2)$$

for individual breeding bulls of Pied Cattle. Notation used in this model is the following

$y(t)$ . . . . body weight of an animal at age t,

α . . . . asymptote of the growth process (weight of a mature animal),

β . . . . is associated with the birth weight,

γ . . . . characterises growth rate.

Least squares estimates of the model parameters were obtained by the Levenberg-Marquardt algorithm, which is a modification of the Gauss-Newton method. Numerical values of the parameters point estimates of the Gompertz model in parametrization (2) are presented in Tab. 1. In this table $\hat{\alpha} = A, \hat{\beta} = B, \hat{\gamma} = C$. The precision of the estimates is characterized by the asymptotic standard error (ASE) and asymptotic 95% confidence interval.

Along with the asymptotic characteristics, results obtained by the bootstraping are presented: standard error (BSE) and 95% trimmed range for the parameters estimates. In this case, the point estimates of Gompertz model parameters were obtained by the sequential quadratic programming algorithm. The coincidence between the point estimates obtained by the two algorithms was high (although in several cases the convergence by the sequential quadratic programming algorithm was not reached).

Generally, there was no dramatic difference between asymptotic standard error (ASE) of the parameter estimates and its bootstrap alternative (BSE) in most cases, although the BSE was usually larger. This could perhaps be expected as the Gompertz model has a reasonably low degree of nonlinearity for the bulls growth data (see Nešetřilová, 1998) and thus its statistical behaviour is relatively closed to properties of linear regression models. In consequence, the closed to asymptotic properties are shown already by samples of size $n \approx 30$. Nevertheless, there are exceptions for some data sets, as documented also in Tab. 1. (Bulls No. 1, 2).

It can be noted that the 95% trimmed range for the growth model parameters is not only set further appart but in some cases shows asymetry when compared to the asymptotic 95% confidence interval. But again, there are also cases where the bootstrap is more restictive on the value of the point estimate.

The calculations presented here were carried out by SPSS 11.5. The bootstrap option of the nonlinear regression allows also the estimation of correlations between model parameters.

For the considered data, the bootstrap correlation coefficients between pairs of (the Gompertz) model parameters were slightly higher than corresponding elements of the asymptotic correlation matrix, nevertheless, in individual cases the increase of correlation coefficient could reach up to 55%.

Tab. 1. Estimates of the parameters of the Gompertz growth model (2) for breeding bulls of Pied Cattle.

Bull No.	n	Parameter estimates	ASE[1]/ BSE[2]	Asymptotic 95% Confidence Interval 95% Trimmed Range lower	upper
1	28	A = 924	31 / 44	861 829	987 1008
		B = 0,926	,046 /,040	,830 ,852	1,022 ,997
		C = 0,0027	,0002 / ,0003	,0022 ,0022	,0031 ,0032
2	26	*A = 1023*	27 / 33	967 954	1079 1089
		B = 0,931	,043 /,058	,841 ,793	1,020 1,032
		C = 0,0028	,0002 / ,0003	,0024 ,0023	,0032 ,0034
3	32	*A = 927*	28 / 29	869 847	985 964
		B = 0,913	,037 /,036	,834 ,830	,989 ,966
		C = 0,0027	,0002 / ,0002	,0023 ,0024	,0030 ,0031
4	32	*A = 1142*	23 / 24	1094 1091	1190 1191
		B = 0,937	,026 /,027	,885 ,877	,989 ,989
		C = 0,0025	,0001 / ,0001	,0023 ,0023	,0027 ,0028
5	27	*A = 902*	23 / 30	855 818	950 950
		B = 0,927	,056 / ,066	,810 ,787	1,043 1,049
		C = 0,0031	,0002 / ,0003	,0027 ,0026	,0036 ,0039

[1] ASE . . . asymptotic standard error
[2] BSE . . . bootstrap standard error
n . . . . . number of observations

Literature:

NEŠETŘILOVÁ, H., PULKRÁBEK, J.: Použití různých typů růstových funkcí k modelování vývoje hmotnosti českého strakatého skotu. *Živoč. Výr*. **40**, 1995, 245-248.

NEŠETŘILOVÁ, H.: *Růstové modely v chovu skotu.* Czech University of Agriculture, Prague, 1998.

RATKOWSKY, D.A.: *Nonlinear Regression Modeling.* Marcel Dekker, Inc., New York and Basel, 1983.

WILCOX, R.R.: Fundamentals of Modern Statistical Methods. Springer, New York and Berlin, 2001.

Kontakt:

Helena Nešetřilová, Department of Statistics, Czech University of Agriculture, Prague, Kamýcká 129, 165 21 Praha 6 - Suchdol, Czech Republic, tel. (+420) 224 382 253,
e-mail: nesetrilova@pef.czu.cz