A C T A U N I V E R S I T A T I S L O D Z I E N S I S
FOLIA OECONOMICA 286, 2013

Wojciech Zieliński[*]

COMPARISON OF ESTIMATORS OF A PROBABILITY OF SUCCESS IN TWO MODELS

Abstract. In modeling two valued phenomena a binomial or negative binomial model is applied. In the paper minimum variance unbiased estimators of a probability of success obtained in two models are compared.

Key words: estimation of probability of success, binomial model, negative binomial model.

Consider a two-valued phenomena:

$$\text{the outcome} = \begin{cases} 1 \quad (\textit{success}), & \text{with probability } \theta, \\ 0 \quad (\textit{fail}), & \text{with probability } 1-\theta. \end{cases} \tag{1}$$

The problem is in estimation of θ.

There are two methods of providing an experiment.

Method 1. The number of all observations is fixed, say n. In those observations the number of successes is counted. This number is a random variable. Let us denote it by ξ.

Method 2. The observations are collected till the fixed number of successes, say r, is observed. Here the number of zeros is a random variable, which will be denoted by η.

Those two models will be compared due to the precision of estimation of probability θ as well as due to the costs of the experiment. In comparison the minimal variance unbiased estimators will be employed, but similar results may be obtained for maximum likelihood and Bayes estimators.

Random variable ξ has a binomial distribution $Bin(n,\theta)$ with probability distribution function

$$f_\theta(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}, \quad x = 0,1,\ldots,n. \tag{2}$$

[*]Professor, Department of Econometrics and Statistics, Warsaw University of Life Sciences and Department of the Prevention of Environmental Hazards and Allergology, Medical University of Warsaw.

The cumulative distribution function (cdf) of ξ may be written as

$$F_\theta(x) = \sum_{i \le x} f_\theta(i) = \beta(n-x, x+1; 1-\theta), \tag{3}$$

where $\beta(a,b;\cdot)$ is a cdf of Beta distribution with parameters (a,b). The statistical model for ξ is as follows:

$$\left(\{0,1,\ldots,n\}, \{Bin(n,\theta), 0 < \theta < 1\}\right). \tag{4}$$

Random variable η has a negative binomial distribution $NB(r,\theta)$ with probability distribution function

$$g_\theta(x) = \binom{r+x-1}{r-1} \theta^r (1-\theta)^x, \quad x = 0,1,2,\ldots. \tag{5}$$

The cdf of η may be written as

$$G_\theta(x) = \sum_{i \le x} g_\theta(i) = \beta(r, x+1; \theta). \tag{6}$$

The statistical model for η is as follows:

$$\left(\{0,1,\ldots\}, \{NB(r,\theta), 0 < \theta < 1\}\right). \tag{7}$$

In the binomial model (4) the unbiased estimator with minimal variance is

$$\hat{\theta}_{MW} = \frac{\xi}{n}. \tag{8}$$

The variance of that estimator equals

$$R_{MW}(\theta) = \frac{\theta(1-\theta)}{n}. \tag{9}$$

The variance of the estimator (for $n = 100$) is shown in Figure 1.

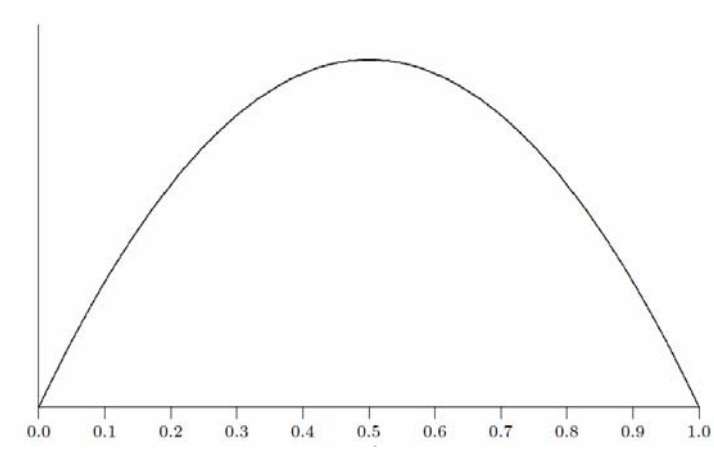

Fig. 1. The variance of the $\hat{\theta}_{MW}$

Note that the variance is symmetric about $\theta = 0.5$ and gains its maximal value at this point.

In the negative binomial model (7) the unbiased estimator with minimal variance is

$$\tilde{\theta}_{MW} = \frac{r-1}{\eta + r - 1}, \tag{10}$$

with the variance

$$R_{MW}(\theta) = \theta^r {}_2F_1(r-1, r-1; r; 1-\theta) - \theta^2. \tag{11}$$

Here ${}_2F_1(a,b;c;x)$ is the hypergeometric function:

$$\begin{gathered} {}_2F_1(a,b;c;x) = \sum_{k=0}^{\infty} \frac{a^{(k)} b^{(k)}}{c^{(k)}} \cdot \frac{x^k}{k!}, \\ a^{(k)} = a(a+1)\cdots(a+k-1), \quad a^{(0)} = 1. \end{gathered} \tag{12}$$

The variance of $\tilde{\theta}_{MW}$ is presented in Figure 2 (for $r = 10$).

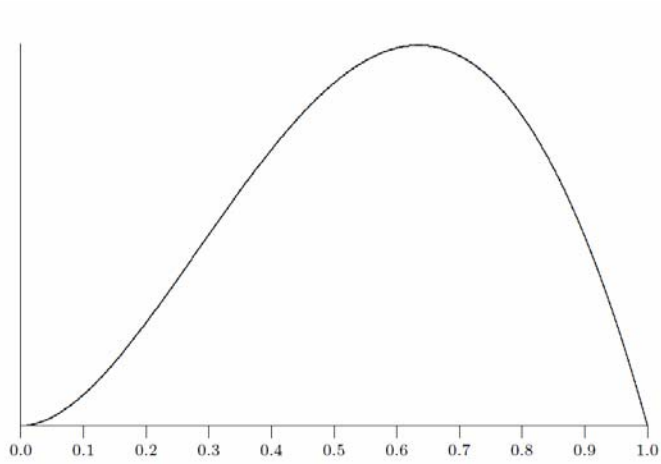

Fig. 2. The variance of the $\widetilde{\theta}_{MW}$

Note that this variance is not symmetric about $\theta = 0.5$.

The question is, what is a minimal sample size to gain given precision of estimation. Let $\varepsilon > 0$ and $\gamma \in (0,1)$ be given numbers. In the model (4) we are looking for minimal n such that

$$P_\theta\left\{\left|\hat{\theta}_{MW} - \theta\right| \le \varepsilon\right\} \ge \gamma, \qquad \forall\, \theta \in (0,1). \tag{13}$$

The above inequality may be written in the form

$$P_\theta\left\{left_{\,\xi}(\theta) \le \xi \le right_{\,\xi}(\theta)\right\} \ge \gamma, \qquad \forall\, \theta \in (0,1), \tag{14}$$

where

$$\begin{aligned} left_{\,\xi}(\theta) &= \min\{\max\{n(\theta - \varepsilon), 0\}, n\}, \\ right_{\,\xi}(\theta) &= \max\{\min\{n(\theta + \varepsilon), n\}, 0\}. \end{aligned} \tag{15}$$

The inequality may be written with the aid of the beta cdf:

$$\begin{aligned} &\beta\left(n - right_{\,\xi}(\theta), right_{\,\xi}(\theta) + 1; 1 - \theta\right) + \\ &- \beta\left(n - left_{\,\xi}(\theta) + 1, left_{\,\xi}(\theta); 1 - \theta\right) \ge \gamma. \end{aligned} \tag{16}$$

This inequality may be solved numerically. The minimal sample sizes are given in the column $n(\theta)$ of Table 1 (for $\varepsilon = 0.01$ and $\gamma = 0.95$). The values of n depend on θ. The maximal value of n is reached for $\theta = 0.5$: at the point at which the variance of $\hat{\theta}_{MW}$ is maximal.

In the model (7) we are looking for minimal r such that

$$P_\theta\left\{\left|\tilde{\theta}_{MW} - \theta\right| \le \varepsilon\right\} \ge \gamma, \qquad \forall\, \theta \in (0,1). \tag{17}$$

This inequality may be written as

$$P_\theta\left\{left_\eta(\theta) \le \eta \le right_\eta(\theta)\right\} \ge \gamma, \qquad \forall\, \theta \in (0,1), \tag{18}$$

where

$$\begin{aligned} \text{left}_\eta(\theta) &= \max\left\{(r-1)\left(\frac{1}{\min\{\theta+\varepsilon,1\}} - 1\right), 0\right\}, \\ \text{right}_\eta(\theta) &= \begin{cases} (r-1)\left(\dfrac{1}{\max\{\theta-\varepsilon,0\}} - 1\right), & \text{for } \theta > \varepsilon, \\ +\infty, & \text{for } \theta \le \varepsilon. \end{cases} \end{aligned} \tag{19}$$

The inequality may be written with the aid of the beta cdf:

$$\beta\left(r, right_\eta(\theta) + 1; \theta\right) - \beta\left(r, left_\eta(\theta); \theta\right) \ge \gamma. \tag{20}$$

As in the (4) model this inequality may be solved numerically. The minimal values of minimal number of successes are given in the column $r(\theta)$ of Table 1. In the column $r(\theta) + E_\theta\eta$ there is given an expected length of the experiment in the (7) model. The values of r depend on θ. The maximal value of $r(\theta)$ is reached at the point at which the variance of $\hat{\theta}_{MW}$ is maximal.

There arise two questions:

1. what is a probability that negative binomial experiment will be shorter than binomial one;

2. what are the costs of experiments in both models?

The answer to the first question may be obtained by calculating the probability

$$P_\theta\{r(\theta) + \eta \le n(\theta)\} = \beta(r(\theta), n(\theta) - r(\theta) + 1; \theta). \tag{21}$$

The values of that probability are given in Table 1 in the column *P*. It is seen that for θ less that 0.8 this probability is quite big. It means that we have very big chances to draw smaller amount of experimental units in negative binomial scheme than in binomial one. For larger values of θ this probability is very small, and for values close to 1 it is zero: minimal sample size in the (4) model is less that number r of required successes in the (7) model.

To answer the second question the cost of the single experiment must be given. Assume that the cost of the single experiment equals one ECU. In the binomial model, the overall cost equals the number $n(\theta)$ (for $\theta = 0.02$ it is 634 ECU).

Table 1. Comparison of models: $\varepsilon = 0.01$, $\gamma = 0.95$

θ	$n(\theta)$	$r(\theta)$	$r(\theta)+E_\theta\eta$	P	cost
0.02	634	16	800	0.20683	–143.02
0.03	1007	35	1167	0.24136	–99.50
0.04	1367	60	1500	0.27279	–58.10
0.05	1719	92	1840	0.26614	–42.31
0.06	2062	131	2183	0.30765	–1.73
0.07	2397	176	2514	0.26622	–7.15
0.08	2724	227	2838	0.27020	11.76
0.09	3043	284	3156	0.26914	26.41
0.10	3355	347	3470	0.26186	36.39
0.20	6045	1229	6145	0.26469	143.17
0.30	7967	2417	8057	0.26141	191.69
0.40	9119	3683	9208	0.23033	187.43
0.50	9503	4795	9590	0.19163	153.41
0.60	9119	5514	9190	0.18757	112.73
0.70	7967	5617	8024	0.17085	61.48
0.80	6045	4869	6086	0.14793	17.76
0.90	3355	3044	3382	0.08250	–11.11
0.91	3043	2828	3108	0.00007	–51.66
0.92	2724	2550	2772	0.00079	–37.37
0.93	2397	2256	2426	0.01589	–20.82
0.94	2062	1994	2121	0.00086	–29.54
0.95	1719	1678	1766	0.00000	–43.29
0.96	1367	1351	1407	0.00012	–23.67
0.97	1007	1058	1091	0.00000	–37.40
0.98	634	681	695	0.00000	–59.20

The expected cost of the whole experiment in the negative binomial model, for given θ, is

$$\sum_{x=left_{\eta}(\theta)}^{right_{\eta}(\theta)} x\binom{r(\theta)+x-1}{r(\theta)-1}\theta^{r(\theta)}(1-\theta)^{x}. \tag{22}$$

For example, for $\theta = 0.02$ we have ($\varepsilon = 0.1$, $\gamma = 0.95$)

$$left_{\eta}(0.02) = 485 \text{ and } right_{\eta}(0.02) = 1485. \tag{23}$$

The expected cost of the whole experiment is ($r(0.02) = 16$)

$$\sum_{x=485}^{1485} x\binom{15+x}{15}0.02^{16}0.98^{x} \approx 777.02. \tag{24}$$

Hence, for $\theta = 0.02$, the binomial model is cheaper than the negative binomial one at about 143.02 ECU.

For other values of θ the differences in costs between the binomial and the negative binomial model are shown in the last column of Table 1. For θ about zero or one the binomial model is cheaper than the negative binomial model. For other values of θ the cheaper is the negative binomial model.

In applications the value of θ is not known and before setting the experiment one should decide which model is to be involved. As a criterion the mean (with respect to θ) cost of experiment may be considered:

$$\int_0^1 \left(n(\theta) - \sum_{x=left_{\eta}(\theta)}^{right_{\eta}(\theta)} x\binom{r(\theta)+x-1}{r(\theta)-1}\theta^{r(\theta)}(1-\theta)^{x} \right) d\theta. \tag{25}$$

Calculations for $\varepsilon = 0.01$ and $\gamma = 0.95$ show that the average difference in costs is about 280 ECU. It means, that expected cost of the experiment in the negative binomial model is smaller than in the binomial one.

In a similar way other estimators may be compared. In general, the estimators of θ in the (4) model and in the (7) model, respectively, are

$$\hat{\theta} = \frac{\xi + a}{n + b}, \qquad \tilde{\theta} = \frac{r + c}{\eta + r + d}, \tag{26}$$

where a,b,c,d are known constants. Unbiased minimal variance estimators are defined by $a=b=0$ and $c=d=-1$; Maximum Likelihood estimators are: $a=b=c=d=0$; Bayesian (with a priori beta with parameters α and β): $a=c=\alpha$, $b=d=\alpha+\beta$. For such estimators

$$\begin{aligned} left_{\xi}(\theta) &= \min\{\max\{(n+b)(\theta-\varepsilon)-a,0\},n\}, \\ right_{\xi}(\theta) &= \max\{\min\{(n+b)(\theta+\varepsilon)-a,n\},0\}, \end{aligned} \tag{27}$$

and

$$\begin{aligned} left_{\eta}(\theta) &= \max\left\{\frac{r+c}{\min\{\theta+\varepsilon,1\}}-(r+d),0\right\}, \\ right_{\eta}(\theta) &= \begin{cases} \dfrac{r+c}{\max\{\theta-\varepsilon,0\}}-(r+d), & \theta>\varepsilon, \\ +\infty, & \theta\le\varepsilon. \end{cases} \end{aligned} \tag{28}$$

Appropriate calculations may be done with the aid of a mathematical software.

REFERENCES

Information on binomial and negative binomial distributions may be found in all textbooks on elementary probability. The mentioned estimators may be found in textbooks on mathematical statistics. In what follows the exemplary textbooks are given.

Bartoszewicz, J. (1996) *Wykłady ze statystyki matematycznej*, wyd. II, PWN Warszawa.

Bartoszyński R., Niewiadomska-Bugaj M. (1996) *Probability and statistical inference*. Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, Inc., New York.

Collani von E., Dräger K. (2001) *Binomial distribution handbook for scientists and engineers*. Birkhuser Boston, Inc., Boston, MA.

Newcomb R. G. (2012) *Confidence intervals for proportions and related measures of effect size*, Chapman & Hall.

Zieliński R. (1990) *Siedem wykładów wprowadzających do statystyki matematycznej*, PWN, Warszawa.

Zieliński R. (2008) *Estymacja frakcji*, Matematyka Stosowana 9(50), 76-90.

Zieliński W. (2010) *Estymacja wskaźnika struktury*, Wydawnictwa SGGW, Warszawa.

Wojciech Zieliński

PORÓWNANIE ESTYMATORÓW PRAWDOPODOBIEŃSTWA SUKCESU W DWÓCH MODELACH

Do modelowania zjawisk dychotomicznych wykorzystuje się model dwumianowy lub model ujemny dwumianowy. W pracy porównano estymatory nieobciążone o minimalnej wariancji prawdopodobieństwa sukcesu w tych dwóch modelach.