GOODNESS OF FIT WITH THE HELP OF A SHARPENED BERNOULLI INEQUALITY

B. BARABÁS and L. SEBESTYÉN

Department of Civil Engineering Mathematics, Technical University, H-1521 Budapest
Received December 28, 1984
Presented by Prof. Dr. J. Reimann

Abstract

In this paper it is intended to apply the inequality

$$\frac{\varepsilon^2}{\frac{9}{4}x_{i-1}^2} \geqq P(|\xi| \geqq |x_i|)$$

to test the examined probability variable ζ has a distribution function $F(x)$.

It is intended to test the hypothesis that the observed probability variable ζ has distribution function $F(x)$.

The sample concerning probability variable ζ having element n yielded by n number of experiments be denoted by $x_1, x_2, \dots x_n$; while the corresponding ordered statistics by $x_1^*, x_2^*, \dots x_n^*$. On the basis of this sample the empirical distribution function $F_n(x)$ of probability variable ζ is determined

$$F_n(x) = \begin{cases} 0, & \text{if } x \leqq x_1^* \\ \frac{k}{n}, & \text{if } x_k^* < x \leqq x_{k+1}^* \\ 1, & \text{if } x_n^* < x. \end{cases} \tag{1}$$

The value of the empirical distribution function $F_n(x)$ of probability variable ζ in case of arbitrary x is a probability variable whose possible values are number $0, \frac{1}{n}, \frac{2}{n}, \dots \frac{n}{n}$ and the corresponding probabilities, i.e. the probability distribution of probability variable $\xi = F_n(x)$

$$P\left(\zeta = \frac{k}{n}\right) = \binom{n}{k} [F(x)]^k [1 - F(x)]^{n-k}, \quad (k = 0, 1, \dots n). \tag{2}$$

Namely, the value of $F_n(x)$ in the arbitrary x point is $\frac{k}{n}$, if the interval $(-\infty, x)$ contains exactly k element, of the sample on the other hand, the probability of the event that ζ has a value smaller than x, is $P(\xi < x) = F(x)$. Accordingly, probability variable $n\xi = nF_n(x)$ $(-\infty < x < +\infty)$ is a probability variable of bionominal distribution having parameters n, $p = F(x)$. Considering this fact its expected value is

$$m(\xi) = \frac{1}{n} E\big(n\ F_n(x)\big) = \frac{1}{n} n F(x) = F(x), \tag{3}$$

and the variance

$$\sigma^2 = \frac{1}{n^2} n F(x)\,\big(1 - F(x)\big) = \frac{1}{n} F(x)\,\big(1 - F(x)\big). \tag{4}$$

If ξ is a probability variable of binomial distribution, then in case of arbitrary positive ε we have

$$P(|\zeta - m| \geqq \varepsilon) \leqq \frac{\sigma^2}{\frac{9}{4}\varepsilon^2}, \tag{5}$$

where m is the expected value of probability variable ξ, and σ^2 is the variance, therefore, here

$$P(|F_n(x) - F(x)| \geqq \varepsilon) \leqq \frac{F(x)\,(1 - F(x))}{\frac{9}{4} n \cdot \varepsilon^2}. \tag{6}$$

Applying the inequality to the complementary event

$$P(|F_n(x) - F(x)| < \varepsilon) > 1 - \frac{F(x)\,(1 - F(x))}{\frac{9}{4} n\,\varepsilon^2}. \tag{7}$$

This statement is proved in the paper of L. Sebestyén entitled "The Chebishev inequality in case of probability variable of special distribution function".

According to the above mentioned, in case of the assumed distribution function $F(x)$ and of certainty $1 - \alpha$ for ε the following critical value is obtained in case of arbitrary x

$$\varepsilon_{kr}^{(\alpha)} = \frac{2}{3}\sqrt{\frac{F(x)\,\big(1 - F(x)\big)}{n\,\alpha}} = \frac{1}{3\sqrt{\alpha}\sqrt{n}}\sqrt{4F(x)\,(1 - F(x))} \tag{8}$$

Thereafter, the hypothesis is accepted at a probability level $1 - \alpha$ if the deviation of the assumed distribution function $F(x)$ from the empirical distri-

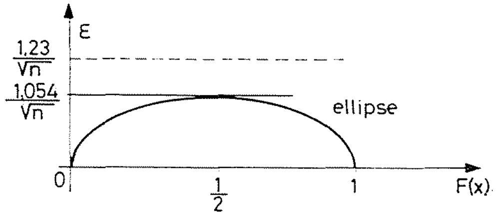

Fig. 1

bution function $F_n(x)$ is smaller than $\varepsilon_{kr}^{(\alpha)}$, in an opposite case this hypothesis is rejected. On the basis of equation (8) the number of experiments needed to the prescribed accuracy ε, as well as to the similarly prescribed certainty $1 - \alpha$ level can be determined in advance.

Finally, the above mentioned probe has been compared to the Kolmogorov—Smirnov probe falling nearest.

Be e.g. the certainty-level of $1 - \alpha$: 90% and the number of experiments n. In this case $\alpha = 0.1$ and

$$\varepsilon_{kr}^{(0.1)} = \frac{1}{3\sqrt{0.1}\sqrt{n}}\sqrt{4F(x)\,(1 - F(x))} = \frac{1.054}{\sqrt{n}}\sqrt{4F(x)\,(1 - F(x))} \qquad (9)$$

From this equation it can be seen that $\varepsilon_{kr}^{\alpha}$ is depend on x and $F(x)$, resp. and the highest deviation allowed is presented on Fig. 1.

In the case of the same level and accuracy, the Kolmogorov—Smirnov probe allows a maximal deviation $\frac{1.23}{\sqrt{n}}$ at any value of x. In Fig. 1 this has been marked by broken line.

If the certainty level is 95%, i.e. $\alpha = 0.05$, then

$$\varepsilon_{kr}^{(0.05)} = \frac{1.39}{\sqrt{n}}\sqrt{4F(x)(1 - F(x))}\,, \qquad (10)$$

and the highest deviation allowed by the Kolmogorov—Smirnov probe is $\frac{1.36}{\sqrt{n}}$. Comparison of the two test is given in Fig. 2.

Alfred Rényi, in his book entitled "Probability theory" remarks: the Kolmogorov—Smirnov probe investigates the deviation of $F_n(x)$ from $F(x)$ independently of the value of $F(x)$; i.e. the same importance has been attributed to deviation $|F_n(x) - F(x)| = 0.01$ in a point x_1 where $F(x_1) = 0.5$, i.e. the deviation 0.01 means a relative deviation of 2% and in a point x_2, where $F(x_2) = 0.01$, i.e. where the absolute deviation 0.01 means a relative deviation of

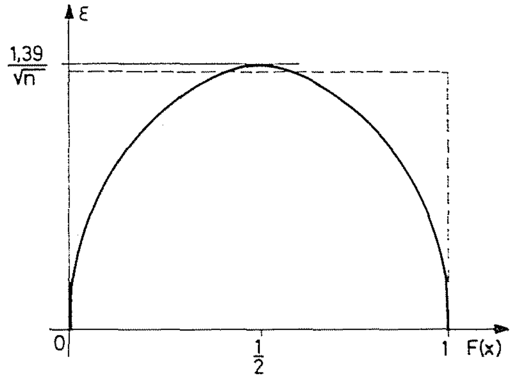

Fig. 2

100%. Therefore, the boundary distribution of probability variable

$$\sqrt{n}\ \sup \left| \frac{F_n(x) - F(x)}{F(x)} \right| \tag{11}$$

is determined and this is the basis of the so-called Kolmogorov—Smirnov—Rényi probe which is an improved version of the Kolmogorov—Smirnov probe.

Now, our probe is compared to that of Kolmogorov—Smirnov—Rényi.

The certainty level $1 - \alpha$ be 90% and the number of experiments n. As already known, the critical value based on the probe suggested by us is

$$\varepsilon_{kr}^{(0.1)} = \frac{1.054}{\sqrt{n}} \sqrt{4F(x)\,(1 - F(x))}\ .$$

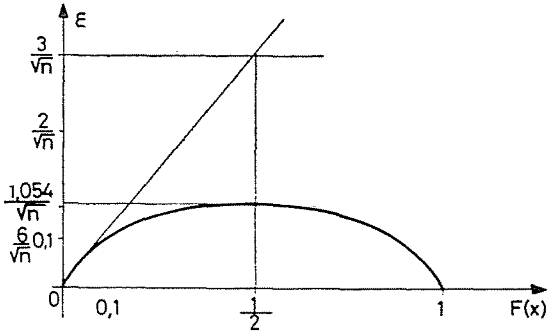

Fig. 3

At the same time, using the Kolmogorov—Smirnov—Rényi probe and involving exclusively those x values for which $F(x) \geqq 0.1$, the highest deviation allowed is

$$\frac{6}{\sqrt{n}} F(x)$$

The certainty range corresponding to both probes is presented in Fig. 3.

The certainty range of Kolmogorov—Smirnov—Rényi probe is demonstrated by the area below the straight line, while that of our probe is to be seen as the area under the ellipse.

References

1. PREKOPA, A.: Probability Theory with Technical Application.* Műszaki Könyvkiadó, Budapest, 1972.
2. RÉNYI, A.: Theory of Probability.* Tankönyvkiadó, Budapest, 1954.

Béla BARABÁS
Dr. Lukács SEBESTYÉN
H-1521 Budapest

* In Hungarian.