Publ. Math. Debrecen
76/3 (2010), 341–346

Note on the Cramér-von Mises test with estimated parameters

By GENNADI MARTYNOV (Moscow)

Dedicated to the 100th anniversary of the birthday of Béla Gyires

Abstract. The asymptotic distribution of the parametric Cramér-von Mises statistic depends on an unknown parameter. In 1955 it was stated (see [4]) that this dependence is absent for the distribution family with the location and scale parameters. We present here the second class of the parametric distribution families with such a property. This is the family with the power and scale parameters.

1. Introduction

This paper investigates the asymptotic distribution of the Cramér-von Mises statistic related to the Weibull and Pareto distribution with estimated parameters. Let $X^n = \{X_1, X_2, \dots, X_n\}$ be the sample from the r.v. with the distribution function $F(x)$, $x \in R_1$. We will test the hypothesis

$$H_0 : F(x) \in \mathbf{G} = \{G(x,\theta),\ \theta = (\theta_1, \theta_2, \dots \theta_k)^\top \in \Theta \subset R_k\},$$

where θ is an unknown vector of parameters. The Cramér-von Mises statistic for testing H_0 is

$$\omega_n^2 = n \int_{-\infty}^{\infty} (F_n(x) - G(x,\theta_n))^2 \, dG(x,\theta_n),$$

Mathematics Subject Classification: 62G10, 62G30.
Key words and phrases: hypotheses testing, goodness-of-fit, Cramér-von Mises test, parameters estimations, Weibull distribution, Pareto distribution.
This paper is supported by Russian foundation for fundamental research: 09-01-00740-a.

where θ_n is an estimator of θ and $F_n(x)$ is the empirical distribution function. The exact methods for calculating the limit distribution are developed mostly for the Cramér-von Mises statistic (see [7], [8]).

The general theory of parametric goodness-of-fit tests based on the empirical process has been developed in [4]. Let θ_n be the maximum likelihood estimator of θ. Under the certain number of the regularity conditions and under H_0, the limit distribution of the statistic ω_n^2 coincides with the distribution of the functional

$$\omega^2 = \int_0^1 \xi^2(t, \theta_0) dt$$

of the Gaussian process $\xi(t, \theta_0)$ with $E\xi(t, \theta_0) = 0$, and covariance function

$$K(t, \tau) = E(\xi(t, \theta_0)\xi(\tau, \theta_0)) = K_0(t, \tau) - q^\top(t, \theta_0) I^{-1}(\theta_0) q(\tau, \theta_0).$$

Here $K_0(t, \tau) = \min(t, \tau) - t\tau$, $t, \tau \in (0, 1)$, θ_0 is a true but unknown value of the parameter θ,

$$q^\top(t, \theta) = (\partial G(x, \theta)/\partial \theta_1, \ldots, \partial G(x, \theta)/\partial \theta_k)|_{t=G(x,\theta)},$$

and $I(\theta)$ is the Fisher information matrix,

$$I(\theta) = (E((\partial/\partial\theta_i) \log g(X, \theta)(\partial/\partial\theta_j) \log g(X, \theta)))_{1 \le i,j \le k}\,,$$
$$g(x, \theta) = \partial G(x, \theta)/\partial x.$$

The distribution of ω^2 depends generally from θ_0 and the distribution family $\mathbf{G}$. Khmaladze [5] has proposed the method of empirical process transformation for eliminate such a dependance. Khmaladze and Haywood [6] has applied this method to exponentiality testing by the Cramér-von Mises statistic.

We will consider here the traditional approach. It is well known that the empirical process does not depend on unknown parameter θ_0 for the distribution family of the form

$$\mathbf{G} = \{G((x - m)/\sigma),\ -\infty < x < \infty,\ \sigma > 0\}.$$

The most known example of such family is the normal distribution family (see [3], [4]). We will propose here another class of the distribution family

$$\mathbf{R} = \{R((x/\beta)^\alpha),\ \alpha > 0,\ \beta > 0,\ x \in \mathbf{X} \subset [0, \infty)\}$$

with this property, where $\mathbf{X}$ is the support of the distribution $R((x/\beta)^\alpha)$. Here $R(z)$ is a distribution function with a corresponding support $\mathbf{Z}$. Particular cases of such families are Weibull and Pareto distributions.

2. General result

Let $X^n = \{X_1, X_2, \ldots, X_n\}$ be the sample from the random variable with a distribution function $F(x),\ x \in R_1$. We will test the hypothesis

$$H_0 : F(x) \in \mathbf{R} = \{R((x/\beta)^\alpha)),\ \alpha > 0,\ \beta > 0,\ x \in \mathbf{X} \subset [0, \infty)\},$$

where α and β are unknown parameters. The set of the alternative distributions contains all another distributions. Here $R(z)$ is the distribution function with a support $\mathbf{Z}$. We note the corresponding density function by $r(z)$. The Cramér-von Mises and Kolmogorov–Smirnov tests are based on the empirical process $\xi_n(x) = \sqrt{n}(F_n(x) - R((x/\hat{\beta})^{\hat{\alpha}})))$, where $\hat{\alpha}$ and $\hat{\beta}$ are here the ML estimates of α and β. Let the regularity conditions are fulfilled. Then we can write the following covariance function for the transformed to $(0, 1)$ limit Gaussian process $\xi(t)$ by formulas from the Section 1:

$$K(t, \tau) = \min(t, \tau) - t\tau - (1/(B_{11}B_{22} - B_{12}^2))$$
$$\times(B_{22}s_1(t)s_1(\tau) - B_{12}(s_1(t)s_2(\tau) + s_2(t)s_1(\tau)) + B_{11}s_2(t)s_2(\tau)),\ t, \tau \in (0,\ 1),$$

$$B_{11} = \int_{\mathbf{Z}} \left(\frac{z \log z\, r'(z)}{r(z)} + \log z + 1 \right)^2 r(z)dz, \quad B_{22} = \int_{\mathbf{Z}} \left(\frac{z\, r'(z)}{r(z)} + 1 \right)^2 r(z)dz,$$

$$B_{12} = \int_{\mathbf{Z}} \left(\frac{z \log z\, r'(z)}{r(z)} + \log z + 1 \right) \left(\frac{z\, r'(z)}{r(z)} + 1 \right) r(z)dz,$$

and

$$s_1(t) = r(R^{-1}(t))R^{-1}(t) \log(R^{-1}(t)), \quad s_2(t) = r(R^{-1}(t))R^{-1}(t).$$

It follows from these formulas that the limit distributions of the considered statistics do not depend on the parameters α and β. Let β be known. Then the covariance function of the process $\xi(t)$ is following:

$$K(t, \tau) = \min(t, \tau) - t\tau - s_1(t)s_1(\tau)/B_{11}.$$

3. Connection between families G and R

Let X be random variable with the distribution $R((z/\beta)^\alpha)$. We can transform X to the another random variable W as follows: $W = -\log(X)$. Then

$$P(W < x) = 1 - R\left(\left(\frac{e^{-x}}{\beta} \right)^\alpha \right) = 1 - R\left(e^{-\frac{x + \log \beta}{1/\alpha}} \right) = G\left(\frac{x - m}{\sigma} \right),$$

where $G(x) = 1 - R(e^{-x})$, and a new parameters of the family $\mathbf{G}$ are connected with the parameters of the family $\mathbf{R}$ by the formulas $m = -\log\beta$, $\sigma = 1/\alpha$. This transformation for the Weibull distribution was considered in [1], [9] and [10]. Inverse transformation from a family $\mathbf{G}$ to a family $\mathbf{R}$ is $X = \exp(-W)$. For example, the normal family changes to the reparametrized lognormal distribution. For convenience sake, the transformations $W = \log(X)$ and $X = \exp(W)$ can also be used.

4. Pareto distribution

We will consider the Pareto distribution in the form

$$F(x) = 1 - (x/\beta)^{-\alpha}, \quad x \geq \beta \geq 0, \quad \alpha > 0.$$

For this distribution $R(z) = 1 - 1/z$ and $\mathbf{Z} = [1, \infty]$. It exists the supereffective unbiased estimate of β

$$\hat{\beta} = \frac{n\alpha - 1}{n\alpha} \min_{i=1,\ldots,n} X_i.$$

We can transform the sample $X_1, \ldots, X_n$ to new sample $Y_1, \ldots, Y_n$, where $Y_i = X_i/\hat{\beta}$. The limit process $\xi(t)$ is equivalent to the process with $\beta = 1$. The MLE of parameter α is

$$\hat{\alpha} = n \Big/ \sum_{i=1}^{n} \log X_i.$$

Hence the covariance function of $\xi(t)$ is

$$K(t,\tau) = \min(t,\tau) - t\tau - (1-t)\log(1-t))(1-\tau)\log(1-\tau)$$

and

$$s_1(t) = -(1-t)\log(1-t), \quad B_{11} = 1.$$

This covariation function coincides with the corresponding covariance function for the exponential family

$$F(x) = 1 - \exp(-x/\beta), \quad \beta \geq 0,\ x \geq 0.$$

We note additionally, that the Pareto family transforms by the transformation $W = \log X$ to the distribution family $1 - e^{\alpha x}$, $0 < x < \infty$. The exponential family belongs to both type of families $\mathbf{G}$ and $\mathbf{R}$. Independence of limit distribution of the statistics ω_n^2 for Pareto family was noted in [2].

5. Weibull distribution

Consider the Weibull distribution family with two parameters

$$F(x) = 1 - e^{-(x/\beta)^{-\alpha}}, \quad x \geq 0,\ \beta \geq 0,\ \alpha > 0.$$

We can note that $R(z) = 1 - e^{-z}$ and $\mathbf{Z} = [0, \infty]$. Maximum likelihood estimates $\hat{\beta}$ and $\hat{\alpha}$ for β and α can be found by numerical methods from the equation system

$$\hat{\beta} = \left(\frac{1}{n} \sum_{i=1}^{n} X_i^{\hat{\alpha}} \right)^{1/\hat{\alpha}}, \frac{n}{\hat{\alpha}} + \log \left(\frac{X_1 \cdot \cdots \cdot X_n}{\hat{\beta}^n} \right) - \sum_{i=1}^{n} \left(\frac{X_i}{\hat{\beta}} \right)^{\hat{\alpha}} \log \left(\frac{X_i}{\hat{\beta}} \right) = 0.$$

The covariance function of $\xi(t)$ in this example has the following elements (see [10]):

$$\begin{aligned}
s_1(t) &= -(1-t)\log(1-t)\log(-\log(1-t)), \\
s_2(t) &= -(1-t)\log(1-t), \\
B_{11}(t) &= \int_0^\infty ((1-z)\log\ z - 1)^2\ e^{-z} dz = (1-C)^2 + \frac{\pi^2}{6}, \\
B_{12}(t) &= \int_0^\infty ((1-z)\log\ z - 1)(1-z)\ e^{-z} dz = 1 - C, \\
B_{22}(t) &= \int_0^\infty (1-z)^2\ e^{-z} dz = 1, \\
B_{11}B_{22} - B_{12} &= \pi^2/6,
\end{aligned}$$

where C is the Euler constant.

The Weibull family transforms by the logarithmic transformation to the extreme value distribution (see [1], [9]).

6. Power distribution on $[0, 1]$

We consider now the distribution function

$$F(x) = \left(\frac{x-a}{b-a} \right)^{\alpha}, \quad x \in [a, b],\ b > a,\ \alpha > 0.$$

Supereffective estimates exist for the parameters a and b. Hence, we can transform the sample to the interval $[0, 1]$ without changing the limit distribution of the statistics. It is sufficient to consider tests for the hypothetical distribution family

$$F(x) = x^{\alpha}, \quad x \in [0, 1],\ \alpha > 0,$$

with $R(z) = z$, $\mathbf{Z} = [0, 1]$. It's easy to derive the covariance function of the limit empirical process $\xi(t)$:

$$K(t, \tau) = \min(t, \tau) - t\tau - t \log t \, \tau \log \tau.$$

The power distribution on $[0, 1]$ can be transformed by the logarithmic transformation to the exponential distribution. The limit distribution of ω_n^2 for this distribution coincides with the corresponding statistics distributions for the exponential and Pareto distribution and for the Weibull distribution with known parameter α. Corresponding tables was found in [9] by simulation.

References

[1] M. Chandra, N. D. Singpurvalla and M. Stephens, Kolmogorov statistics for tests of fit for the extreme-value and Weibull distributions, *Journal of American Statistical Association* **76** (1981).

[2] V. Choulakian and M. A. Stephens, Goodness-of-fit tests for the generalized Pareto distribution, *Technometrics* **43** (2001), 478–484.

[3] I. I. Gikhman, One conception from the theory of ω^2-test, *Nauk. Zap. Kiiv Univ.* **13** (1954), 51–60 (in *Urainian*).

[4] M. Kac, J. Kiefer and J. Wolfowitz, On tests of normality and other tests of goodness-of-fit based on distance methods, **30** (1955), 420–447.

[5] E. V. Khmaladze, A martingale approach in the theory of parametric goodness-of-fit tests, *Theor. Prob. Appl.* **26** (1981), 240–257.

[6] E. Khmaladze and J. Haywood, On distribution-free goodness-of-fit testing of exponentiality, *J. Econometrics* **143** (2008), 5–18.

[7] G. V. Martynov, The Omega Square Tests, *Nauka*, *Moscow*, 1979 (in *Russian*).

[8] G. V. Martynov, Statistical tests based on empirical processes and related questions, *J. Soviet. Math.* **61** (1992), 2195–2271.

[9] M. Stephens, Tests Based on EDF Statistics, in Goodness-of-fit Techniques, (R. D'Agostino and M. Stephens, eds.), *Marcel Dekkers*, *New York*, 1986, 97–193.

[10] Yu. Tyurin and N. E. Savvushkina, An agreement criterion for the Weibull–Gnedenko distribution, *Izv. Akad. Nauk SSSR Tekhn. Kibernet*, no. 3 (1984), 109–112 (in *Russian*).

GENNADI MARTYNOV
INSTITUTE FOR INFORMATION TRANSMISSION PROBLEMS
OF THE RUSSIAN ACADEMY OF SCIENCES
(KHARKEVICH INSTITUTE
BOLSHOY KARETNY PER. 19
MOSCOW, 127994
RUSSIA

E-mail: martynov@iitp.ru
URL: http://www.guem.iitp.ru

(Received June 6, 2009; revised December 22, 2009)