G. Martynov (IITP RAS) **Cramér-von Mises test for parametric dictribution family** [1]

0. Introduction. Let $X^n = \{X_1, X_2, ..., X_n\}$ be the sample from the r.v. with the distribution function $F(x)$, $x \in R_1$. We will test the hypothesis $H_0: \quad F(x) \in \mathcal{G} = \{G(x,\theta),\ \theta \in R_k\}$, where θ is an unknown vector of parameters. We will consider the Cramér-von Mises statistic

$$\omega_n^2(\hat{\theta}_n) = n \int_{-\infty}^{\infty} (F_n(x) - G(x,\hat{\theta}_n))^2\, dG(x,\hat{\theta}_n) \equiv n \int_0^1 (\hat{F}_n(t) - t)^2 dt,$$

where $\hat{\theta}_n$ is the maximum likelihood estimator of θ, $F_n(x)$ and $\hat{F}_n(t)$ are the empirical distribution functions, based on the samples X^n and $T^n = G(X^n, \hat{\theta}_n)$, correspondingly. Under some regularity conditions,

$$\omega_n^2(\hat{\theta}_n) \ \to_d \ \omega^2(\theta^0) = \int_0^1 \xi^2(t,\theta^0) dt, \tag{1}$$

where $\xi(t,\theta^0)$ is the Gaussian process with zero mean and with some covariance function $K(t,\tau,\theta^0)$. It follows that in the general case, the distribution of the Cramér-von Mises statistic may depend on all unknown parameters or on their part.

It is well known that the empirical process does not depend on unknown parameter θ^0 for the family of the form (see [1, 4])

$$\mathcal{G} = \{G((x-\theta_1)/\theta_2),\ -\infty < x < \infty,\ \theta_2 > 0\}.$$

Another class of the distribution family proposed in [6] is

$$\mathcal{R} = \{R((x/\beta)^\alpha),\ \alpha > 0,\ \beta > 0, x \in [0,\infty)\}.$$

Both of these families are closely interconnected. Here we want to consider the two cases when the distribution of the $\omega^2(\theta^0)$ depends on the parameters. One of the methods for calculating the asymptotic distributions of the statistics under consideration is also discussed.The alternative method for testing such hypotheses is set out in [5].

1. Gamma distribution family. Asymptotic distribution of the Cramér-von Mises statistic for the gamma distribution family

$$G(x;\theta,\kappa) = \frac{\Gamma(\kappa, x/\theta)}{\Gamma(\kappa)} \equiv H\left(\frac{x}{\theta}, \kappa\right),\ -\infty < x < \infty,\ \theta > 0,\ \kappa > 0,$$

depends on one unknown parameter κ. This follows from the theorem below.

[1] e-mail: martynov@iitp.ru, magevl@gmail.com

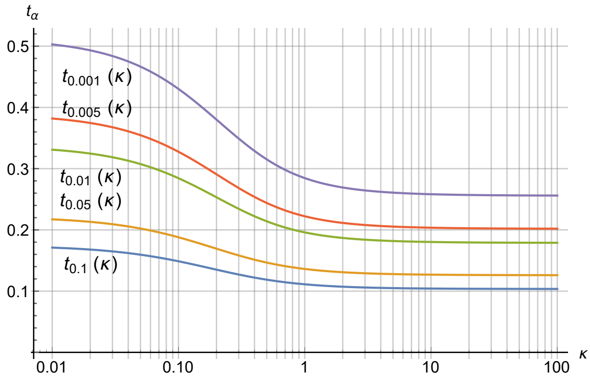

Figure 1. Asymptotic upper critical levels of the Cramér-von Mises statistic for gamma distribution family

Theorem 1. *The covariance function of the corresponding to the gamma distribution family asymptotic empirical process can be represented as follow*

$$K(t,\tau;\kappa) = \min(t,\tau) - t\tau - C(\kappa)\big(e^{-u}u^{\kappa}W(v,\kappa) + \kappa W(u,\kappa)W(v,\kappa)$$
$$+e^{-v}v^{\kappa}W(u,\kappa) - e^{-(u+v)}(uv)^{\kappa}\psi'(\kappa)\big)\Big|_{u=H^{-1}(t,\kappa),\ v=H^{-1}(\tau,\kappa)},$$
$$C(\kappa) = 1/(\Gamma(\kappa)^2(\kappa\psi'(\kappa)-1)),$$
$$W(z,\kappa) = \frac{z^{\kappa}}{\kappa^2}\,{}_2F_2(\{\kappa,\kappa\},\{1+\kappa,1+\kappa\};-z) + (\Gamma(\kappa,z)-\Gamma(\kappa))(\ln z - \psi(\kappa)),$$

${}_2F_2(\cdot)$ is a generalized hypergeometric function, $\Gamma(\kappa,z)$ is the upper incomplete gamma function, $\psi(z)$ is the digamma function.

A detailed five-digit table was calculated, on the basis of which table 1 was drawn. The independence of the distribution of statistics in question in this section of κ was briefly noted in [7].)

2. Family of exponentiated distribution function. The exponentiated exponential distribution $F(x;t,\kappa) = (1-e^{x/t})^{\kappa},\ \kappa > 0,\ t > 0$ was introduced by [3]. This definition can be generalized to the form exponentiated distribution function

$$F(x;t,\kappa) = G^{\kappa}\left(\frac{x}{t}\right),\ \kappa > 0,\ t > 0,$$

where $G(\cdot)$ is a continuous distribution function. The asymptotic distribution of the Cramér-von Mises statistic for such a family depends generally on two parameters.

3. Approximation of the integral of a squared Gaussian process. Here we present an effective method of calculating the distributions of integrals from squared Gaussian processes [2]. The integral in (1) will be approximated by the sum as follows

$$\omega^2 = \int_0^1 \xi^2(t)dt \approx \sum_{i=1}^{m} \xi^2(t_i),$$

where $\xi(t)$ is the Gaussian process with zero mean and a covariance function $K(t,\tau)$. Here, we are interested in the Darboux sums, when $a_i = 1/m$. We will use in the sequence the following notation:

$$A(t,\tau) = 2K^2(t,\tau), \;\; t_i = \tau_i = (i-1/2)/m,$$
$$\underline{A}^{k,l}(\tau) = \lim_{t\uparrow\tau} \frac{\partial^{k+l}}{\partial t^k \partial \tau^l} A(t,\tau), \;\; \overline{A}^{k,l}(\tau) = \lim_{t\downarrow\tau} \frac{\partial^{k+l}}{\partial t^k \partial \tau^l} A(t,\tau).$$

Theorem 2. *Let there exist all derivatives up to forth order from the function $A(t,\tau)$ on $\{0 \le t,\tau \le 1, t \ne \tau\}$ and the derivatives are uniformly bounded on this square without the diagonal. Then the variance $D\varepsilon_m$ can be represented asymptotically in the form*

$$D\varepsilon_m = \frac{c_2}{m^2} + O\left(\frac{1}{m^4}\right), \;\; c_2 = \frac{1}{12}\int_0^1 \left(\underline{A}^{1,0}(t) - \overline{A}^{1,0}(t)\right) dt.$$

This approximation was applied to the calculation of the above table. First, the eigenvalues $\lambda_i, \; i = 1,\dots,m$ of the matrix $(K(t_{i,j}, \;\; i,j \in (1,\dots m))$ are calculated. Then these values are substituted in the Smirnov formula to calculate the distributions of quadratic forms from normal r.v. If you use the eigenvalues of the covariation function $K(t,\tau)$, then the correction of the quadratic form will be required and the Smirnov formula can not be used.

BIBLIOGRAPHY

[1] Anderson, T. W., Darling, D. A. Asymptotic theory of certain "Goodness of Fit" criteria based on stochastic processes. The Ann. of Math. Statist. 1952. V. 23. No. 2. P. 193 –212.

[2] Deheuvels, P., Martynov G. Cramer-von-Mises-type tests with applications to test of independence for multivariate extreme-value distributions. Communications in Statistics-Theory and Methods. 1996. V. 25. No. 4. P. 871 – 908.

[3] Gupta, R. D. and Kundu, D. (1999). Theory & methods: Generalized exponential distributions. Australian & New Zealand Journal of Statistics, 41(2), 173–188.

[4] Kac, M., Kiefer, J., Wolfowitz, J. On tests of normality and other tests of goodness-of-fit based on distance methods. Ann. Math. Statist. 1955. V. 30. P. 420 –447.

[5] Khmaladze E.V. A martingale approach in the theory of parametric goodness-of-fit tests. Theor. Prob. Appl. 1981. 26. No 2. P. 240 – 257.

[6] Martynov G. Note on the Cramer-von Mises test with estimated parameters. Publ. Math. Debrezen. 2010. V. 76. No. 3. P. 341 – 346.

[7] Stephens M. A. Tests based on EDF statistics. In D'Agostino R. B., Stephens M. A., eds. Goodness-of-Fit Techniques, Marcel Dekker, New York/Basel, 1986. P. 97 – 193.