UDC 519.2

# Anderson-Darling and New Weighted Cramér-von Mises Statistics

**G. V. Martynov***

* *Institute for Information Transmission Problems of the RAS, Bolshoy Karetny per., 19, build.1, Moscow, 127051, Russia*

**Abstract.** Anderson-Darling statistic modifies the classical empirical process in the interval $[0, 1]$ by multiplying it by a weighting function $\psi(t) = (t(1-t))^{-1/2}$. The weighting function redistributes the test sensitivity to deviations of the alternative distribution function from the hypothetical one between different subsets of [0,1]. In practice, the tests can be of interest with other weighting functions. There a new formulas was proposed for eigenfunctions of the Anderson-Darling statistics. Also, it was analyzed a statistic "inverse" to the Anderson-Darling statistic with the weighting function $\psi(t) = (t(1-t))^{1/2}$. The theory is based on the use of various special functions book [1]. In practice, could be useful the Cramér-von-Mises tests with the weighting functions, belonging to family $\psi(t) = t^{\alpha}(1-t)^{\beta}$, $\alpha > -1$, $\beta > -1$. The paper contains a table of distribution for statistics with different values of the degrees $\alpha > -1$ and $\beta > -1$. The table was calculated without using the statistical simulation.

**Keywords:** Cramér-von-Mises test, Anderson-Darling statistic, goodness-of-fit test, weighting function, eigenvalues, eigenfunctions, statistical tables.

## 1. Introduction: Weighted Cramér-von Mises test

One-dimensional weighted Cramér-von Mises statistic is

$$\omega_n^2 = n \int_0^1 \psi^2(t)(F_n(t) - t)^2 dt,$$

where $F_n(t)$ is the empirical distribution function based on the observations $X_1, X_2, ..., X_n$ of the uniformly distributed on $[0, 1]$ random variable, and $\psi(t)$ is a weighting function. The statistic (1) designed to test the hypothesis

$$H_0: \quad F(t) = t, \quad \text{against the alternative } H_1: \quad F(t) \neq t, \quad t \in [0, 1],$$

where $F(x)$ is continuous distribution function.

If the condition

$$\int_0^1 \psi^2(t)\, t\, (1-t) dt < \infty$$

is fulfilled then the statistic $\omega_n^2$ converges in probability to

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

where $\xi(t),\ t \in [0,1]$, is the Gaussian process with zero mean and the covariance function

$$K_\psi(t,\tau) = \psi(t)\psi(\tau)(\min(t,\tau) - t\tau).$$

The Gaussian process $\xi(t)$ can be developed in the Karhunen-Loève series

$$\xi(t) = \sum_{i=1}^{\infty} \frac{x_k \varphi_k(t)}{\sqrt{\lambda_k}},$$

where $x_k,\ k = 1,2,...$, are the independent standard normal random variables, and $\lambda_k$ and $\varphi_k(t)$, $i = 1,2,...$, are the eigenvalues and eigenfunctions of the linear operator with the kernel $K(t,\tau)$, i.e. solutions of the Fredholm integral equation

$$\varphi(t) = \lambda \int_0^1 \psi(t)\psi(\tau)(\min(t,\tau) - t\tau)\varphi(\tau)d\tau.$$

Under the contigual alternatives $H_1 : F_n(t) = t + \delta(t)/\sqrt{n},\ n = 1,2,...$, the distribution of $\omega^2$ is the distribution of noncentral quadratic form

$$Q = \sum_{i=1}^{\infty} \frac{(x_k + \delta_k)^2}{\lambda_k}, \text{ where } \delta_k = \int_0^1 \delta(t)\varphi_k(t)dt.$$

By twice differentiation (2) respect to $t$ , we obtain differential equation

$$h''(t) + \lambda\psi^2(t)h(t) = 0$$

with the conditions $h(0) = h(1) = 0$. Here, $h(t) = \varphi(t)/\psi(t)$.

Deheuvels and Martynov in article [3] described for $\psi(t) = t^\beta$ the follows result. Let $\{B(t) : 0 \le t \le 1\}$ be the Brownian bridge. Then, for each $\beta = \frac{1}{2\nu} - 1 > -1$, the Karhunen-Loeve expansions of $\{\xi(t) = t^\beta B(t) : 0 < t \le 1\}$ is given by

$$t^\beta B(t) = \sum_{k=1}^{\infty} \frac{x_k e_{kB}(t)}{\sqrt{\lambda_{kB}}}, \quad e_{B,k}(t) = \frac{t^{\frac{1}{2\nu}-\frac{1}{2}} J_\nu\left(z_{\nu,k}\, t^{\frac{1}{2\nu}}\right)}{\sqrt{\nu} J_{\nu-1}\left(z_{\nu,k}\right)}, \quad 0 < t \le 1.$$

Here, $\{\omega_k : k \ge 1\}$ are i.i.d. $N(0,1)$ random variables, and, for $k = 1,2,\ldots$, the eigenvalues are $\lambda_k = (z_{\nu,k}/2\nu)^2$, $z_{\nu,k},\quad k = 1,2,...$, are zeros of the Bessel functions $J_\nu(z)$.

It is considered also the Cramér-von Mises statistic of the form

$$\omega_n^2(a,b) = n \int_0^1 t^{2a}(1-t)^{2b}(F_n(t) - t)^2 dt$$

with the weighting function $\psi(t) = t^a(1-t)^b, a > -1, \;\; b > -1$. The well known case is the Anderson-Darling statistic

$$A_n^2 = \omega_n^2(-0.5, -0.5) = n \int_0^1 \frac{(F_n(t) - t)^2}{t(1-t)} dt$$

with $a = -0.5$ and $b = -0.5$. For the statistic $A_n^2$ equation (3) is transformed to

$$t(1-t)h''(t) + \lambda h(t) = 0, \; h(0) = h(1) = 0.$$

Anderson and Darling in article [2] found that their statistic has $\lambda_k = k(k+1)$ and $h_k(t) = \sqrt{t(1-t)} P_k'(2t-1), \; k = 1, 2, ...,$ where $P_k(t), \; k = 1, 2, ...,$ are the Legendre polynomials. The information related to the subject of this work can also be found in book [4, 5, 5, 6].

## 2. New formulas for eigenfunctions for the the Anderson-Darling statistic

At first, we will propose a direct method for deriving expressions for considering eigenfunctions. We will find the possible solutions of the differential equation (4) in the form

$$h_k(t) = t(1-t)(1 + a_{1,\,k}t + a_{2,\,k}t^2 + ... + a_{k-1,\,k}t^{k-1}) \equiv t(1-t)V_k(t), \;\; k = 1, 2, ...$$

Here, $h_1(t)$ and $h_2(t)$ are understood to be $h_1(t) = t(1-t)$ and $h_2(t) = t(1-t)(1 + a_{1,\,2}t)$. This solution satisfy to the conditions $h(0) = h(1) = 0$.

**Theorem 1** *The solutions of the equation (4) can be represented for each* $\lambda = \lambda_k = k(k+1), \; k = 1, 2, ...,$ *as*

$$V_k(t) = 1 + \beta_{1,k}t + \beta_{1,k}\beta_{2,k}t^2 + \beta_{1,k}\beta_{2,k}\beta_{3,k}t^3 + ... + \beta_{1,k}...\beta_{k-1,k}t^{k-1}$$

*or*

$$V_k(t) = ((((\beta_{k-1,\,k}t + 1)\beta_{k-2,\,k}t + ... + 1)\beta_{3,\,k}t + 1)\beta_{2,\,k}t + 1)\beta_{1,\,k}t + 1,$$

*where* $\beta_{s,\,k} = 1 - \lambda_k/\lambda_s = 1 - \frac{k(k+1)}{s(s+1)}, \; s = 1, 2, ..., k-1.$

The following theorem solution can be represents another solution with using of the hypergeometric functions.

**Theorem 2** *The solutions of the equation (4) can be represented for each* $\lambda = \lambda_k = k(k+1), \; k = 1, 2, ...,$ *as*

$$h_k(t) = t \cdot {}_2F_1(-k, \; k+1; \; 2; \; t), \;\; k = 1, 2, ...$$

This result can be derived from the fact that the equation (4) is particular case of the equation for hypergeometric function ${}_2F_1$.

**Theorem 3** *The following identity is valid:*

$$(1-t)V_k(t) \equiv {}_2F_1(-k,\ k+1;\ 2;\ t).$$

**Theorem 4** *The normalized eigenfunctions of the covariance operator corresponding to the Anderson-Darling statistic can be written as:*

$$\varphi_k(t) = 2\sqrt{k(k+1)(2k+1)}\sqrt{t(1-t)}V_k(t)$$

$$= 2\sqrt{k(k+1)(2k+1)}\sqrt{\frac{t}{(1-t)}}\ {}_2F_1(-k,\ k+1;\ 2;\ t) =$$

$$= 2\sqrt{\frac{2r+1}{k(k+1)}}\sqrt{t(1-t)}P_k'(2t-1),$$

$$k = 1, 2, ...,\quad t \in [0,1].$$

## 3. Statistic with $\psi(t) = \sqrt{t(1-t)}$

Here, we will consider the statistic Cramér-von Mises with $\psi(t) = \sqrt{t(1-t)}$. It is "inverse" for Anderson-Darling statistic The equation (3) have the form $h''(t) + \lambda t(1-t)h(t) = 0$ with the conditions $h(0) = 0$ and $h(1) = 0$. Its solution is

$$h(t) = C_1 \cdot {}_1F_1\left(\frac{1}{4} - \frac{\sqrt{a}}{16}, \frac{1}{2}, \frac{\sqrt{a}(-1+2x)^2}{4}\right)\exp(\sqrt{a}x(1-x)/2)$$

$$+C_2 \cdot {}_1F_1\left(\frac{3}{4} - \frac{\sqrt{a}}{16}, \frac{3}{2}, \frac{\sqrt{a}(-1+2x)^2}{4}\right)(-1+2x)\exp(\sqrt{a}x(1-x)/2).$$

Here, ${}_1F_1(a;b;z)$ is the Kummer confluent hypergeometric function. With applying the conditions $h(0) = h(1) = 0$, the following equations can be derived

$$h(0) = C_1 \cdot {}_1F_1\left(\frac{1}{4} - \frac{\sqrt{a}}{16}, \frac{1}{2}, \frac{\sqrt{a}}{4}\right) - C_2 \cdot {}_1F_1\left(\frac{3}{4} - \frac{\sqrt{a}}{16}, \frac{3}{2}, \frac{\sqrt{a}}{4}\right) = 0,$$

$$h(1) = C_1 \cdot {}_1F_1\left(\frac{1}{4} - \frac{\sqrt{a}}{16}, \frac{1}{2}, \frac{\sqrt{a}}{4}\right) + C_2 \cdot {}_1F_1\left(\frac{3}{4} - \frac{\sqrt{a}}{16}, \frac{3}{2}, \frac{\sqrt{a}}{4}\right) = 0.$$

Hence, the equation for eigenvalues is the determinant of the previous equation

$$ {}_1F_1\left(\frac{1}{4}-\frac{\sqrt{a}}{16},\frac{1}{2},\frac{\sqrt{a}}{4}\right)\cdot {}_1F_1\left(\frac{3}{4}-\frac{\sqrt{a}}{16},\frac{3}{2},\frac{\sqrt{a}}{4}\right)=0. $$

This equation can be resolved by numerical methods. First zeros of the left multiplier are:
45.24420524 514.3565172 1495.598597 2988.879509 4994.177419
7511.484585 10540.79747 14082.11420 18135.43367 22700.75517
First zeros of the right multiplier are:
215.7714658 940.9700448 2178.236123 3927.526974 6188.830121
8962.140456 12247.45544 16044.77364 20354.09420 25175.41653

## 4. Another weighted statistics

The equation (3) with $\psi(t) = (t(1-t))^{3/2}$ has the solution with Heun triconfluent function. The same equation with $\psi^2(t) = 1-\cos(\pi t)$ has the solution with Mathieu function.

## 5. Table of the distribution for statistics with $\psi(t) = (t(1-t))^{\alpha}$

In the Table 1 we present the quantiles of the $\omega^2(\alpha, \beta$ distribution with $\alpha = \beta$. More detailed tables can be found in article [7].

Table 1

Quantiles of $\omega^2(\alpha,\alpha)/S$

| $p/\alpha$ | – 0.70 | – 0.50 | – 0.25 |
|---|---|---|---|
| 0.90 | 4.3776 | 1.9330 | 0.7945 |
| 0.95 | 5.4914 | 2.4924 | 1.0437 |
| 0.99 | 8.2378 | 3.8781 | 1.6606 |
| 0.995 | 9.4680 | 4.4982 | 1.9362 |
| $S$ | 1 | 1 | 1 |

| 0 | 0.5 | 1 | 2 |
|---|---|---|---|
| 3.4731 | 0.7213 | 1.5834 | 0.8245 |
| 4.6136 | 0.9709 | 2.1479 | 1.1284 |
| 7.4346 | 1.5872 | 3.5397 | 1.8769 |
| 8.6939 | 1.8620 | 4.1603 | 2.2105 |
| 10 | 10 | 100 | 1000 |


## Acknowledgments

The work was carried out at IITP RAS and supported by Russian Science Foundation (grant RSF No. 14-50-00150).


## References


1. *Abramowitz M., Stegun I. A.* Handbook of Mathematical Function. — National Bureau of Standards. Applied Nathematics Series. — 1972. — Vol. 55.
2. *Anderson T. W., Darling D. A.* Asymptotic theory of certain "Goodness of Fit" criteria based on stochastic processes // The Annals of Mathematical Statistics. — 1952. — Vol. 23, no. 2. — P. 193–212.
3. *Deheuvels P., Martynov G.* Karhunen-Loève expansions for weighted Wiener processes and Brownian bridges via Bessel functions // Progress in Probability. — Birkhäuser, Basel/Switzerland, 2003. — Vol. 55. — P. 57–93.
4. *Durbin J., Knott M.* Components of the Cramér-von-Mises statistics. I. // J. R. Statist. Soc., Ser. B, — 1972. — Vol. 34. — P. 290–307.
5. *Martynov G. V.* The Omega Square Tests. — M.: Nauka, 1978. (in Russian).
6. *Martynov G. V.* Statistical tests based on empirical processes and related questions // J. Soviet. Math. — 1992. — Vol. 61. — P. 2195–2271. (Original Russian Text published in Progress in Science and Technology, Series on Probability Theory, Mathematical Statistics, Theoretical Cybernetics (Itogi Nauki i Tekhniki), — VINITI Acad. of Sc. USSR, Moscow, 1990. — Vol. 30. — P. 3–112.)
7. *Martynov G. V.* Anderson–Darling statistic and its "inverse" // Journal of Communications Technology and Electronics. —2016. — Vol. 61, no. 6. — P. 709–716. ( Original Russian Text published in Informatsionnye Protsessy. — 2015. — Vol. 15, no. 3. — P. 377–388).