

Статистика Андерсона–Дарлинга и “обратная” к ней¹

Г.В. Мартынов

*Институт проблем передачи информации им. А. А. Харкевича РАН, г. Москва, Россия,
Высшая школа экономики, Москва, Россия
email: martynov@iitp.ru*

Поступила в редколлегию 14.09.2015

Аннотация—Более шестидесяти лет критерий Андерсона–Дарлинга остается наиболее применяемым среди всех критериев Крамера–Мизеса (омега-квадрат). В статистике критерия Андерсона–Дарлинга классический эмпирический процесс, заданный на интервале $[0, 1]$ умножается на весовую функцию $\psi(t) = (t(1-t))^{-1/2}$. Это действие перераспределяет чувствительность критерия к отклонениям функции распределения наблюдаемой случайной величины от гипотетической функции распределения на различных её участках. Однако для практики могут представлять интерес и критерии с другими весовыми функциями. Работа предоставляет новые формулы для собственных значений статистики Андерсона–Дарлинга. Рассматривается также статистика, “обратная” к статистике Андерсона–Дарлинга с весовой функцией $\psi(t) = (t(1-t))^{1/2}$. При применении взвешенных статистик Крамера–Мизеса могут представлять интерес также критерии с другими весовыми функциями. В настоящей работе представлены таблица квантилей распределения статистик с весовыми функциями вида $\psi(t) = t^\alpha(1-t)^\beta$, $\alpha > -1$, $\beta > -1$. Представленные квантили даны для 36 различных комбинаций параметров $\alpha > -1$ и $\beta > -1$. Таблица вычислена с помощью точных численных методов без использования методов моделирования.

КЛЮЧЕВЫЕ СЛОВА: Критерий Крамера–Мизеса, статистика Андерсона–Дарлинга, критерии согласия, весовые функции, собственные значения, собственные функции.

1. ВВЕДЕНИЕ: ВЗВЕШЕННАЯ СТАТИСТИКА КРАМЕРА-МИЗЕСА

Весовые функции для критериев Крамера–Мизеса были впервые предложены Андерсоном и Дарлингом в 1952-м году. Во введении приводятся некоторые широко используемые результаты из этой работы, необходимые далее. Упомянем также работы Дарбина [1973], Дарбина и Нотта [1972], Дэувэлса и Мартынова [2003] и Мартынова [1979, 1992].

А. Пусть $X_1, X_2, \dots, X_n$ являются независимыми наблюдениями случайной величины X с неизвестной функцией распределения $\tilde{G}(x)$, $x \in (-\infty, \infty)$. Проверяется простая гипотеза

$$\tilde{H}_0 : \tilde{G}(x) = F(x)$$

против альтернативы

$$\tilde{H}_1 : \tilde{G}(x) \neq F(x).$$

Здесь $F(x)$ является непрерывной функцией распределения. Наблюдения X_i , $i = 1, 2, \dots, n$, трансформируем с помощью преобразования $T_i = F(X_i)$ к значениям T_i , $i = 1, 2, \dots, n$. Следовательно, если гипотеза $\tilde{H}$ выполняется, то эти значения могут рассматриваться как наблюдения случайной величины $T = U[0, 1]$, равномерно распределенной на интервале $[0, 1]$. Теперь

¹ Работа выполнена в ИППИ. Разделы 2 и 3 поддержаны грантом Российского Научного Фонда (Проект РНФ 14-50-00150).

мы можем проверить следующую эквивалентную гипотезу

$$H_0 : G(t) = U(t)$$

против альтернативы

$$H_1 : G(x) \neq U(t).$$

В. Для проверки гипотезы H_0 при альтернативе H_1 может использоваться статистика Крамера-Мизеса. Её вариант с весовой функцией $\psi^2(t)$ есть

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

Здесь, $F_n(t)$ является эмпирической функцией распределения, построенная по преобразованным наблюдениям T_i , $i = 1, 2, \dots, n$. Классическая статистика Крамера-Мизеса имеет весовую функции $\psi^2(t) = 1$. Статистика Андерсона-Дарлинга содержит весовую функцию $\psi^2(t) = 1/(t(1-t))$. В этой работе мы рассматриваем также статистику с весовой функцией $\psi^2(t) = (t(1-t))$. В общем случае мы предполагаем, что весовая функция является непрерывной. Следующая теорема дает условия сходимости распределения статистики ω_n^2 к предельному распределению.

Теорема 1. *Условие*

$$\int_0^1 K(t, t) dt < \infty$$

является необходимым и достаточным для сходимости по распределению статистики ω_n^2 к функционалу

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

где $\xi(t)$ является гауссовским процессом с нулевым математическим ожиданием и ковариационной функцией

$$K(t, \tau) = \psi(t)\psi(\tau)(\min(t, \tau) - t\tau), \quad t, \tau \in [0, 1]. \quad (1)$$

Доказательство. Теорема следует немедленно из представления эмпирического процесса

$$\xi_n(t) = \sqrt{n}\psi(t)(F_n(t) - t)$$

в виде нормированной суммы

$$\xi_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [\psi(t)(I_{t \leq T_i} - t)]$$

независимых одинаково распределённых случайных процессов. Эта сумма слабо сходится в гильбертовом пространстве к гауссовскому процессу с ковариационной функцией (1) (см., например, Van der Vaart and Wellner, 1996). $\square$

С. Предельный процесс может быть представлен как $\xi(t) = \psi(t)B(t)$, $t \in [0, 1]$, где $B(t)$ является броуновским мостом с ковариационной функцией

$$K_B(t, \tau) = (\min(t, \tau) - t\tau), \quad t, \tau \in [0, 1].$$

Разложение Карунена–Лозва процесса $\xi(t)$ имеет вид

$$\xi(t) = \sum_{i=1}^{\infty} \frac{x_k \varphi_k(t)}{\sqrt{\lambda_k}}, \quad (2)$$

где x_k , $k = 1, 2, \dots$, являются независимыми стандартными нормально распределенными случайными величинами, λ_k и $\varphi_k(t)$, $i = 1, 2, \dots$, есть собственные значения и собственные функции линейного интегрального оператора с ядром $K(t, \tau)$, то есть решениями интегрального уравнения Фредгольма

$$\varphi(t) = \lambda \int_0^1 \psi(t) \psi(\tau) (\min(t, \tau) - t\tau) \varphi(\tau) d\tau. \quad (3)$$

Дифференцируя дважды (3) по t , мы получаем дифференциальное уравнение

$$h''(t) + \lambda \psi^2(t) h(t) = 0 \quad (4)$$

с условиями

$$h(0) = h(1) = 0.$$

Здесь

$$h(t) = \varphi(t)/\psi(t). \quad (5)$$

Пусть λ_k и $h_k(t)$, $k = 1, 2, \dots$, являются решениями этого уравнения. Тогда значения λ_k являются также собственными значениями уравнения Фредгольма (3), а $\varphi_k(t) = h_k(t)\psi(t)$ его собственными функциями.

Из разложения (2) следует, что распределение ω^2 совпадает при нулевой гипотезе с распределением квадратичной формы

$$Q(t) = \sum_{i=1}^{\infty} \frac{x_k^2}{\lambda_k}.$$

При сближающихся альтернативах $H_1 : F_n(t) = t + \delta(t)/\sqrt{n}$ распределение статистики Крамера–Мизеса совпадает с распределением нецентральной квадратичной формы

$$Q = \sum_{i=1}^{\infty} \frac{(x_k + \delta_k)^2}{\lambda_k},$$

где

$$\delta_k = \int_0^1 \delta(t) \varphi_k(t) dt.$$

Таким образом, для нахождения мощности критерия необходимо знать собственные функции ковариационного оператора.

D. Статистика Крамера–Мизеса с весовой функцией $\psi(t) = t^\beta$ рассмотрена в работе Деувэла и Мартынова (2003). В ней получено разложение Карунена–Лозва предельного процесса $\{t^\beta B(t) : 0 < t \leq 1\}$ при $\beta > -1$

$$t^\beta B(t) = \sum_{k=1}^{\infty} \sqrt{\lambda_{kB}} \omega_k e_{kB}(t).$$

Здесь $\{\omega_k : k \geq 1\}$ – независимые одинаково распределенные случайные величины со стандартным нормальным распределением $N(0, 1)$, $\lambda_k = (2\nu/z_{\nu,k})^2$,

$$e_k(t) = \frac{t^{\frac{1}{2\nu}-\frac{1}{2}} J_\nu \left(z_{\nu,k} t^{\frac{1}{2\nu}} \right)}{\sqrt{\nu} J_{\nu-1}(z_{\nu,k})}, \quad 0 < t \leq 1,$$

и $z_{\nu,k}$, $k = 1, 2, \dots$, являются нулями функции Бесселя $J_\nu(z)$, $\nu = 1/(2(\beta + 1))$.

Е. Для статистики Андерсона-Дарлингга

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

уравнение (4) принимает вид

$$t(1-t)h''(t) + \lambda h(t) = 0, \quad h(0) = h(1) = 0. \quad (6)$$

Андерсон и Дарлинг (1952) нашли, что $\lambda_k = k(k+1)$. Предельное распределение этой статистики является распределением квадратичной формы

$$Q = \sum_{i=1}^{\infty} \frac{x_k^2}{k(k+1)}.$$

Её математическое ожидание есть

$$EA^2 = \int_0^1 K(t, t) dt = EQ = \sum_{i=1}^{\infty} \frac{1}{k(k+1)} = 1. \quad (7)$$

Выражение для нормализованной характеристической функции было получено Андерсоном и Дарлингом (1952), и затем описывалось Дурбиным и Ноттом (1972), Шораком и Веллером (2009), Такехиро Хироцу с соавторами (2011) и в других работах. Эта функция есть

$$2\sqrt{\frac{2k+1}{k(k+1)}} \sqrt{t(1-t)} P_k'^{(2t-1)}, \quad k = 1, 2, \dots, \quad t \in [0, 1].$$

В этой работе мы дадим дополнительно другие выражения для собственных функций статистики Андерсона-Дарлингга, в том числе использующие гипергеометрические функции.

Ф. Вычислена таблица квантилей статистик Крамера-Мизеса с весовыми функциями вида $\psi(t) = t^\alpha(1-t)^\beta$, $\alpha > -1$, $\beta > -1$, для различных комбинаций параметров α и β .

2. СОБСТВЕННЫЕ ФУНКЦИИ ДЛЯ СТАТИСТИКИ АНДЕРСОНА-ДАРЛИНГА

2.1. Прямой метод

Здесь мы рассмотрим непосредственный метод решения уравнения (6) в виде

$$h_k(t) = t(1-t)(1 + a_{1,k}t + a_{2,k}t^2 + \dots + a_{k-1,k}t^{k-1}), \quad k = 1, 2, \dots$$

Здесь, $h_1(t)$ и $h_1(t)$ понимаются как $h_1(t) = t(1-t)$ и $h_2(t) = t(1-t)(1+a_{1,2}t)$, соответственно. Это решение будет удовлетворять условиям $h(0) = h(1) = 0$. Обозначим

$$V_k(t) = 1 + a_{1,k}t + a_{2,k}t^2 + \dots + a_{k-1,k}t^{k-1}, \quad k = 1, 2, \dots$$

Теорема 2. *Решение уравнения*

$$t(1-t)h''(t) + \lambda h(t) = 0 \quad (8)$$

с условиями $h(0) = h(1) = 0$ существует для каждого $\lambda = \lambda_k = k(k+1)$, $k = 1, 2, \dots$, в виде

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

или

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

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

Доказательство. Из (8) следует, что

$$h_k''(t) + \lambda_k V_k(t) = 0. \quad (9)$$

Функция $h_k(t)$ может быть переписана как

$$h_k(t) = t + (a_{1,k} - 1)t^2 + (a_{2,k} - a_{1,k})t^3 + \dots + (a_{k-1,k} - a_{k-2,k})t^k - a_{k-1,k}t^{k+1}.$$

Следовательно,

$$h_k''(t) = 1 \cdot 2(a_{1,k} - 1) + 2 \cdot 3(a_{2,k} - a_{1,k})t + 3 \cdot 4(a_{3,k} - a_{2,k})t^2 + \dots + (k-1) \cdot k(a_{k-1,k} - a_{k-2,k})t^{k-2} - k \cdot (k+1)a_{k-1,k}t^{k-1}.$$

После подстановки $h_k''(t)$ and $V_k(t)$ in (9) получаем равенство

$$[1 \cdot 2(a_{1,k} - 1) + \lambda_k] + [2 \cdot 3(a_{2,k} - a_{1,k}) + \lambda_k a_{1,k}]t + [3 \cdot 4(a_{3,k} - a_{2,k}) + \lambda_k a_{2,k}]t^2 + \dots + [(k-1) \cdot k(a_{k-1,k} - a_{k-2,k}) + \lambda_k a_{k-2,k}]t^{k-2} + [-k \cdot (k+1)a_{k-1,k} + \lambda_k a_{k-1,k}]t^{k-1} \equiv 0.$$

Следовательно, для коэффициентов и λ_k мы имеем следующую систему уравнений:

$$\begin{aligned} 1 \cdot 2(a_{1,k} - 1) + \lambda_k &= 0, \\ 2 \cdot 3(a_{2,k} - a_{1,k}) + \lambda_k a_{1,k} &= 0, \\ 3 \cdot 4(a_{3,k} - a_{2,k}) + \lambda_k a_{2,k} &= 0, \\ &\vdots \\ (k-1) \cdot k(a_{k-1,k} - a_{k-2,k}) + \lambda_k a_{k-2,k} &= 0, \\ -k(k+1)a_{k-1,k} + \lambda_k a_{k-1,k} &= 0. \end{aligned}$$

Если $a_{k-1,k} \neq 0$, то тогда из последнего уравнения следует, что $\lambda_k = k(k+1)$ и

$$a_{1,k} = 1 - \frac{\lambda_k}{\lambda_1}, a_{2,k} = \left(1 - \frac{\lambda_k}{\lambda_2}\right) a_{1,k}, \dots, a_{k-1,k} = \left(1 - \frac{\lambda_k}{\lambda_{k-1}}\right) a_{k-2,k}$$

для всех $k = 1, 2, \dots$. Осюда можно заключить, что действительно $a_{k-1,k} \neq 0$ для всех $k = 1, 2, \dots$. Из (7) следует, что получены все собственные значения и все собственные функции интегрального уравнения Фредгольма.

Пусть $\beta_{s,k} = 1 - \lambda_k/\lambda_s$, $s = 1, 2, \dots, k-1$. Следовательно, для $r < k$ $a_{r,k} = \beta_{1,k}\beta_{2,k} \dots \beta_{r,k}$. Тогда

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

или

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

□

Теорема 3. Собственные значения интегрального уравнения

$$\varphi(t) = \lambda \int_0^1 \frac{\min(t, \tau) - t\tau\varphi(\tau)}{\sqrt{t(1-t)\tau(1-\tau)}} d\tau$$

есть $\lambda_k = k(k+1)$. Соответствующие ненормализованные собственные функции будут

$$\tilde{\varphi}_k(t) = \sqrt{t(1-t)} V_k(t), \quad k = 1, 2, \dots, \quad t \in [0, 1].$$

Доказательство. Результат теоремы следует из равенства (5). $\square$

2.2. Выражение собственных функций статистики Андерсона–Дарлингса через гипергеометрические функции

Пусть ${}_2F_1(a, b; c; t)$ есть гипергеометрическая функция Гаусса, определяемая рядом

$${}_2F_1(a, b; c; z) = 1 + \sum_{m=1}^{\infty} \frac{(a)_m (b)_m}{(c)_m m!} t^m, \quad (10)$$

$(d)_m = d(d+1)\cdots(d+m-1)$, $(d)_0 = 1$, – символ Прохгаммера. Этот ряд сходится для всех $|t| < 1$ при подходящих значениях параметров a , b и c . Функция ${}_2F_1(a, b; c; t)$ является основным решением гипергеометрического дифференциального уравнения

$$t(1-t) \frac{d^2}{dt^2} y(t) + \{c - (a+b+1)t\} \frac{d}{dt} y(t) - a b y(t) = 0. \quad (11)$$

Теорема 4. Уравнение

$$t(1-t)h''(t) + \lambda h(t) = 0 \quad (12)$$

при условиях $h(0) = h(1) = 0$ имеет в качестве решения собственные значения $\lambda_k = k(k+1)$, $k = 1, 2, \dots$, и собственные функции

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

Доказательство. Уравнение (12) может быть рассмотрено как частный случай (11). Параметры a , b and c уравнения (12) удовлетворяют равенствам

$$\begin{cases} c = 0, \\ a + b + 1 = 0, \\ ab = \lambda. \end{cases}$$

Отсюда получаем два решения

$$a_1 = \frac{1 - \sqrt{1+4\lambda}}{2}, \quad b_1 = \frac{1 + \sqrt{1+4\lambda}}{2}, \quad c = 0,$$

и

$$a_2 = \frac{3 - \sqrt{1+4\lambda}}{2}, \quad b_2 = \frac{3 + \sqrt{1+4\lambda}}{2}, \quad c = 0.$$

Для наших целей выберем одно из решений Куммера уравнения (11) (см. 15.5.4 в справочнике Абрамовича и Стеган 1972)

$$h(t) = t^{1-c}(1-t)^{c-a-b} \cdot {}_2F_1((1-a), (1-b; 2-c; t).$$

Это решение удовлетворяет условиям $h(0) = h(1) = 0$. Следует принять во внимание, что в выражении для гипергеометрической функции a и b могут меняться местами. Таким образом, получаем два решения уравнения:

$$h_*(t) = t \cdot {}_2F_1((1 - \sqrt{1 + 4\lambda})/2, (1 + \sqrt{1 + 4\lambda})/2; 2; t),$$

$$h_{**}(t) = t(1 - t) \cdot {}_2F_1((3 - \sqrt{1 + 4\lambda})/2, (3 + \sqrt{1 + 4\lambda})/2; 2; t).$$

Известно, что

$$F(\tilde{a}, \tilde{b}; \tilde{c}; t) = (1 - t)^{\tilde{c} - \tilde{a} - \tilde{b}} F(\tilde{c} - \tilde{a}, \tilde{c} - \tilde{b}; \tilde{c}; t),$$

(формула 15.3.3 из Абрамовича и Стеган, 1972). Из этого следует, что $h_*(t) = h_{**}(t)$. В качестве решения будем использовать выражение $h(t) = h_*(t)$. Таким образом,

$$h(t) = tF(\tilde{a}, \tilde{b}; \tilde{c}; t), \quad (13)$$

где $\tilde{a} = (1 - \sqrt{1 + 4\lambda})/2$, $\tilde{b} = (1 + \sqrt{1 + 4\lambda})/2$ и $\tilde{c} = 2$. Используя известные (и выведенные в предыдущем разделе) значения собственных чисел $\lambda_k = k(k + 1)$, получаем утверждение теоремы. $\square$

Теорема 5. Собственные значения интегрального уравнения

$$\varphi(t) = \lambda \int_0^1 \frac{\min(t, \tau) - t\tau\varphi(\tau)}{\sqrt{t(1-t)\tau(1-\tau)}} d\tau$$

есть $\lambda_k = k(k + 1)$. Соответствующие ненормированные собственные функции есть

$$\tilde{\varphi}_k(t) = \sqrt{\frac{t}{(1-t)}} {}_2F_1(-k, k + 1; 2; t), \quad k = 1, 2, \dots, \quad t \in [0, 1].$$

Доказательство. Результат теоремы следует из равенства (5). $\square$

Теорема 6. Выполняется следующее равенство:

$$(1 - t)V_k(t) \equiv {}_2F_1(-k, k + 1; 2; t). \quad (14)$$

Доказательство. Обе части тождества (14) являются полиномами степени k . Они являются решениями одного и того же уравнения, которые могут отличаться только на константу. Наконец, при $t = 0$ обе функции обращаются в единицу. Отсюда следует утверждение теоремы. $\square$

2.3. Выражение собственных функций через полиномы Лежандра

Формулы для собственных значений ковариационного оператора, связанного со статистикой Андерсона–Дарлинга, выразим теперь через полиномы Лежандра, используя полученные выражения через гипергеометрические функции.

Используем частный случай формулы 15.4.17, приводимой Абрамовичем и Стеган (1972),

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

где $P_k^m(x)$, $m = \pm 1, \pm 2, \dots$, являются ассоциированными полиномами Лежандра или полиномами Феррера. В нашем случае (см. Сансон, 1991, Абрамович и Стеган, 1972)

$$P_k^{-1}(1-2t) = -\frac{\Gamma(k)}{\Gamma(k+2)} P_k^1(2t-1) = -\frac{1}{k(k+1)} P_k^1(2t-1).$$

Так как

$$P_k^1(t) = -\sqrt{1-t^2} \frac{d}{dt} P_k(t),$$

то

$${}_2F_1(-k, k+1; 2; t) = (-1)^k \frac{1-t}{k(1+k)} \frac{d}{dt} P_k(2t-1).$$

Здесь $P_k(t)$ является обычным полиномом Лежандра. Из формулы 8.5.1 в справочнике Абрамовича и Стеган(1972) получаем, что

$$P_k^1(t) = (1-t^2)^{-1/2} (ktP_k(t) - kP_{k-1}(t)).$$

Отсюда получаем, что

$${}_2F_1(-k, k+1; 2; t) = \frac{(-1)^k}{1+k} [(2t-1)P_k(2t-1) - P_{k-1}(2t-1)].$$

Следующая теорема суммирует полученные формулы для характеристической функции.

Теорема 7. Нормализованные собственные функции, связанные со статистикой Андерсона–Дарлингa, могут быть записаны в трех следующих формах

$$\begin{aligned} \varphi_k(t) &= 2\sqrt{k(k+1)(2k+1)}\sqrt{t(1-t)}V_k(t) \\ &= 2\sqrt{k(k+1)(2k+1)}\sqrt{\frac{t}{(1-t)}}{}_2F_1(-k, k+1; 2; t) \\ &= 2\sqrt{\frac{2k+1}{k(k+1)}}\sqrt{t(1-t)}P_k'^{(2t-1)}, \\ &\quad k = 1, 2, \dots, \quad t \in [0, 1]. \end{aligned}$$

Доказательство. Последнее выражение для нормированных собственных функций дано Андерсоном и Дарлингом (1952), Дурбиным и Ноттом (1972), Шораком и Веллнером (2009) и другими авторами. Сравнивая это выражение с другими равенствами из этого раздела, убеждаемся в справедливости теоремы. $\square$

3. СТАТИСТИКА “ОБРАТНАЯ” К СТАТИСТИКЕ АНДЕРСОНА–ДАРЛИНГА

Рассмотрим теперь статистику с весовой функцией $\psi(t) = \sqrt{t(1-t)}$. Эта функция обратна к весовой функции статистики Андерсона–Дарлингa. Статистика имеет вид

$$\bar{A}_n^2 = n \int_0^1 t(1-t)(F_n(t) - t)^2 dt \quad (15)$$

Можно проверить, что

$$h(t) = C_1 \cdot {}_1F_1\left(\frac{1}{4} - \frac{\sqrt{\lambda}}{16}; \frac{1}{2}; \frac{\sqrt{\lambda}(-1+2x)^2}{4}\right) \exp(\sqrt{\lambda}x(1-x)/2) \\ + C_2 \cdot {}_1F_1\left(\frac{3}{4} - \frac{\sqrt{\lambda}}{16}; \frac{3}{2}; \frac{\sqrt{\lambda}(-1+2x)^2}{4}\right) (-1+2x) \exp(\sqrt{\lambda}x(1-x)/2)$$

с константами C_1 и C_2 является решением уравнения, соответствующего статистике (15),

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

Здесь ${}_1F_1(a; b; t)$ – обобщенная гипергеометрическая функция Куммера

$${}_1F_1(a; b; t) = 1 + \sum_{m=1}^{\infty} \frac{(a)_m}{(b)_m m!} t^m.$$

Применив к этому решения условия $h(0) = h(1) = 0$ получаем уравнения

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

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

Отсюда следует, что уравнение для собственных значений получается при приравнении детерминанта предыдущей системы нулю:

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

Это уравнение может быть решено численными методами. Первые 10 корней уравнения

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

есть

45.24420524 514.3565172 1495.598597 2988.879509 4994.177419
7511.484585 10540.79747 14082.11420 18135.43367 22700.75517,

а первые 10 корней уравнения

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

есть

215.7714658 940.9700448 2178.236123 3927.526974 6188.830121
8962.140456 12247.45544 16044.77364 20354.09420 25175.41653.

В приводимых далее таблицах содержится, в частности, и значения квантилей предельного распределения описанной статистики.

4. ТАБЛИЦЫ ПРЕДЕЛЬНЫХ РАСПРЕДЕЛЕНИЙ СТАТИСТИКИ С $\psi^2(t) = t^a(1-t)^b$

Здесь приводится таблица квантилей статистики Крамера–Мизеса

$$\omega_n^2(a, b) = n \int_0^1 t^{2a}(1-t)^{2b}(F_n(t) - t)^2 dt, \quad a > -1, \quad b > -1, \quad (16)$$

при комбинациях значений a и b , $a \leq b$, из множества значений $\{-0.70, -0.50, -0.25, 0, 0.5, 1, 2, 3\}$. Распределение статистики (16) не зависит от перемены местами значений a и b . Для получения квантиля статистики следует значения из таблицы разделить на величину S , приводимую в каждой ячейке.

В клетке $(a, b) = (0, 0)$ таблицы приводятся квантили классической статистики Крамера–Мизеса без весовой функции. В клетке $(-0.5, -0.5)$ приводятся квантили статистики Андерсона–Дарлингга. В клетке $(0.5, 0.5)$ приводятся квантили статистики, “обратной” к статистике Андерсона–Дарлингга с весовой функцией $\psi_2(t) = t(1-t)$. В столбце таблицы с $b = 0$ приводятся значения статистики, исследованной в работе Деувэlsa и Мартынова (2003), с весовой функцией $\psi_2(t) = t^a$.

Все значения квантилей в таблице вычислены численными методами без использования моделирования.

5. ПРИЛОЖЕНИЕ. ВЫЧИСЛЕНИЕ ЗНАЧЕНИЙ СТАТИСТИКИ ω_n^2 ПРИ РАЗЛИЧНЫХ ВЕСОВЫХ ФУНКЦИЯХ

Здесь представлена формула для вычисления значений статистики (16) по заданным наблюдениям $T_1, T_2, \dots, T_n$. Эта формула имеет вид:

$$\omega_n^2(a, b) = nQ_1 + 2 \sum_{j=1}^n \left[V_1(T_j) - \frac{2j-1}{n} V_2(T_j) \right],$$

где

$$Q(t) = \int_0^1 (1-s^2)s^{2a}(1-s)^{2b}ds = \frac{2(2+2a+b)}{3+2a+2b} B(1+2a, 2+2b),$$

$$V_1(t) = \int_0^t s^{2a}(1-s)^{2b}ds = \frac{t^{2+2a}}{2+2a} {}_2F_1(2+2a, -2b, 3+2a, t),$$

$$V_2(t) = \int_0^t s^{2a+1}(1-s)^{2b}ds = \frac{t^{1+2a}}{1+2a} {}_2F_1(1+2a, -2b, 2+2a, t).$$

Здесь, $B(\cdot, \cdot)$ – бета-функция, а ${}_2F_1(\cdot, \cdot; \cdot; \cdot)$ – гипергеометрическая функция.

Эти формулы получены из общей формулы, выведенной Андерсоном и Дарлинггом (1952).

6. БЛАГОДАРНОСТЬ

В заключение, автор выражает благодарность профессору Полю Деувэльсу, (Университет Пьера и Марии Кюри, Париж 6) за полезные обсуждения.

Таблица 1. Квантили взвешенной статистики Крамера–Мизеса

$a \backslash b$	p	– 0.70	– 0.50	– 0.25	0	0.5	1	2	3
– 0.70	0.90	4.3776	3.1073	2.3872	1.9870	1.7467	1.4068	1.004	0.8425
	0.95	5.4914	3.945	3.0504	2.5473	2.3069	1.8474	1.2956	1.0877
	0.99	8.2378	6.0177	4.6918	3.9343	3.6939	2.9385	2.0171	1.6951
	0.995	9.4680	6.9458	5.4267	4.5552	4.3148	3.4268	2.3400	1.9669
	S	1	1	1	1	1	1	1	1
– 0.50	0.90		1.9330	1.3156	0.9976	0.8309	0.5075	3.4030	2.5599
	0.95		2.4924	1.7108	1.3030	1.1364	0.6673	4.4854	3.3779
	0.99		3.8781	2.6900	2.0598	1.8931	1.0633	7.1651	5.4029
	0.995		4.4982	3.1279	2.3980	2.2313	1.2401	8.3620	0.6307
	S		1	1	1	1	1	10	10
– 0.25	0.90			0.7945	0.5470	3.1723	2.1314	1.2020	0.7952
	0.95			1.0437	0.7224	4.2126	2.8383	1.6052	1.0635
	0.99			1.6606	1.1564	6.7855	4.5862	2.6019	1.7264
	0.995			1.9362	1.3503	7.9341	0.5366	3.0466	2.0222
	S			1	1	10	10	10	10
0	0.90				3.4731	1.7667	1.0698	5.1443	3.0161
	0.95				4.6136	2.3614	1.4346	6.9210	4.0645
	0.99				7.4346	3.8312	2.3356	11.307	6.6522
	0.995				8.6939	4.4870	2.7376	13.264	7.8062
	S				10	10	10	100	100
0.50	0.90					0.7213	3.6477	1.3188	3.5014
	0.95					0.9709	4.9284	1.7887	3.9713
	0.99					1.5872	8.0878	2.9473	5.1298
	0.995					1.8620	9.4967	3.4638	5.6464
	S					10	100	100	1000
1	0.90						1.5834	4.4607	1.7162
	0.95						2.1479	6.0764	2.3429
	0.99						3.5397	10.058	3.8871
	0.995						4.1603	11.833	4.5753
	S						100	1000	1000
2	0.90							0.8245	2.2401
	0.95							1.1284	3.0734
	0.99							1.8769	5.1252
	0.995							2.2105	6.0397
	S							1000	10000
3	0.90								4.5257
	0.95								6.2259
	0.99								10.411
	0.995								12.277
	S								100000

Примечание. Для получения квантиля статистики следует значения из таблицы делаться на величину S , приводимую в этой же ячейке.

СПИСОК ЛИТЕРАТУРЫ

1. Abramowitz, M., Stegun, I. A. (1972) Handbook of Mathematical function. National Bureau of Standards. Applied Mathematics Series, **V**. 55.
2. Anderson, T. W., Darling, D. A. (1952) Asymptotic theory of certain "Goodness of Fit" criteria based on stochastic processes. *The Annals of Mathematical Statistics*, **V**. 23, No. 2, pp. 193-212.
3. Deheuvels, P., Martynov, G. (2003) Karhunen-Loève expansions for weighted Wiener processes and Brownian bridges via Bessel functions. *Progress in Probability*., **V**. 55, pp. 57–93. Birkhäuser, Basel/Switzerland.
4. Deheuvels, P., Martynov, G. (2008) A Karhunen-Loeve decomposition of a Gaussian process generated by independent pairs of exponential random variables. *J. Funct. Anal.* **V**. 255, 9, pp. 2363-2394 .
5. Durbin, J. (1973). Distribution theory for tests based upon the sample distribution function. *Regional Conference Series in Applied Mathematics* **V**. 9. S.I.A.M., Philadelphia.
6. Durbin, J., Knott, M. (1972) Components of the Cramér-von-Mises Statistics. I. *J. R. Statist. Soc., B*, **V**. 34 pp. 290-307.
7. Martynov, G. V. (1979). The Omega Square Tests. *Nauka, Moscow*, 80 pp. (in Russian).
8. Martynov, G. V. (1992). Statistical tests based on empirical processes and related questions. *J. Soviet. Math.* **V**. 61, pp. 2195–2271.
9. Sansone, G. (1991). Expansions in Series of Legendre Polynomials and Spherical Harmonics. *Ch. 3 in Orthogonal Functions*, rev. English ed. New York: Dover, pp. 169-294.
10. Shorack, G.R., Wellner, J. A. (1986) Empirical Processes with Applications to Statistics. John Wiley & Sons, 938 pp.
11. Takehiro Hirotsu, Hiroki Kondo, Shingo Saito, Takuya Sato, Tatsushi Tanaka, Setsuo Taniguchi (2011) Anderson-Darling test and the Malliavin calculus. *Mathematics-for-Industry, Ser. A*, **V**. 3, pp. 73-78
12. Van der Vaart, A. W., Wellner, J. A. (1996) Weak Converge and Empirical Processes. Springer-Verlag, 508 pp.

Anderson–Darling Statistic and Its “Inverse”

Martynov G.V.

For more than sixty years, Anderson-Darling statistic is one of the most popular in applications among the Cramér-von-Mises goodness-of-fit tests. This 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 on different subsets of $[0, 1]$. In practice, the tests can be of interest having other weight functions. The paper proposes new formulas for eigenfunctions of the Anderson-Darling statistics. Also, it was analyzed a statistic "inverse" to the Anderson-Darling statistic with the weighting function $\psi(t) = (t(1 - t))^{1/2}$. There was considered also another weighting functions. The proposed theory is based on the use of various special functions. In practice, could be useful the Cramér-von-Mises tests with weighting functions from the family $\psi(t) = t^\alpha(1 - t)^\beta$, $\alpha > -1$, $\beta > -1$. The paper contains a table of distribution for these statistics with different values of the degrees $\alpha > -1$ and $\beta > -1$. The table was calculated by different methods with good precision without using the statistical simulation.

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