Cramér-von Mises Test for Gaussian Processes

Gennady MARTYNOV
Institute for Information Transmission Problems of the Russian Academy of Sciences
&
National Research University Higher School of Economics,
Russia, Moscow

Advanced Finance and Stochastics

June 25, 2013

Weighted Cramér-von Mises statistic

One-dimensional weighted Cramér-von Mises statistic is

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

where $F_n(t)$ is the empirical distribution function based on the sample $X_1, X_2, ..., X_n$ from the uniform distribution on $[0, 1]$, and $\psi(t)$ is a weight function.The statistic (1) designed to test the hypothesis

$$H_0: \quad F(x) = t$$

against the alternative

$$H_1: \quad F(x) \neq t,$$

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 ω_n^2 converges in probability to

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

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

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

(see for example Van der Vaart and Wellner [1996, p. 50]).

The Gauss 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 $x_k \sim N(0,1)$, $k = 1, 2, ...$, are independent random variables, and λ_k and $\varphi_k(t)$, $i = 1, 2, ...$, are the eigenvalues and eigenfunctions of the Fredholm integral equation

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

By twice differentiation (3) with respect to t, we obtain the differential equation

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

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

Classical Cramér-von Mises statistic

This is the statistic

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

with $\psi(t) = 1$. The eigenvalues and eigenfunctions of the covariance function $K(t, \tau)$ are $\lambda_i = (\pi i)^2$ and $\varphi_i(t) = \sqrt{2}\sin(\pi i t)$, $i = 1, 2, \ldots$. The limit in probability ω^2 of ω_n^2 can be written as

$$\omega^2 = \int_0^1 B^2(t)dt = \sum_{k=1}^{\infty} \frac{x_k^2}{(\pi k)^2},$$

where $B(t)$ is the Brown bridge process, i.e. Gaussian process with zero mean and covariance function $K(t, \tau) = \min(t, \tau) - t\tau$.

Cramér-von Mises statistic with the power weight function

Deheuvels and Martynov (2003) described 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 $\{t^{\beta}B(t) : 0 < t \le 1\}$ is given by

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

Here, $\{\omega_k : k \ge 1\}$ are i.i.d. $N(0,1)$ random variables, and, for $k = 1, 2, \ldots$, the eigenvalues are $\lambda_k = (2\nu/z_{\nu,k})^2$, corresponding eigenfunctions are

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

and $z_{\nu,k}$, $k = 1, 2, ...$, are zeros of the Bessel functions $J_{\nu}(z)$.

Eigenvalues for the multivariate uniformity test with weight function

We will use the notations for d-vectors $\mathbf{s} = (s_1, ..., s_d)^\top$ and $\mathbf{t} = (t_1, ..., t_d)^\top$.

Let $\mathbf{U} = (U(1), ..., U(d))^\top$ be a random vector with the uniform distribution function on $[0, 1]^d$

$$F(\mathbf{t}) = P(\mathbf{U} < \mathbf{t}) = \prod_{i=1}^{d} t_i$$

and

$$\mathbf{U}_i = (U_i(1), ..., U_i(d)), \quad i = 1, ..., n.$$

are the n observations of $\mathbf{U}$. The empirical distribution function has a form

$$F_n(\mathbf{t}) = \frac{1}{n} \sum_{i=1}^{n} 1_{\mathbf{U}_i < \mathbf{t}}$$

We can write the multivariate empirical process as

$$\alpha_n(\mathbf{t}) = n^{1/2}(F_n(\mathbf{t}) - F(\mathbf{t})).$$

The classical statistic for the distribution uniformity testing on $[0,1]^d$ is

$$\Omega_n = \int_{[0,1]^d} \alpha_n^2(\mathbf{t}) d\mathbf{t}.$$

We will consider now the weighted variant of this statistic

$$\Omega_n = \int_{[0,1]^d} \mathbf{t}^{2\mathbf{B}} \alpha_n^2(\mathbf{t}) d\mathbf{t},$$

where $\mathbf{t}^{\mathbf{B}} = t_1^{\beta_1} \cdot ... \cdot t_d^{\beta_d}$, $\mathbf{B} = (\beta_1, ..., \beta_d)$.

The weighted process $\mathbf{t}^{\mathbf{B}} \alpha_n(\mathbf{t})$ converges weakly in Hilbert space on $[0,1]^d$ to the Gauss process $\mathcal{X}(\mathbf{t}) = \mathbf{t}^{\mathbf{B}} \mathcal{B}(\mathbf{t})$, where $\mathcal{B}(\mathbf{t})$ is the standard multivariate Brown bridge with the covariation function

$$\mathcal{K}_{\mathcal{B}}(\mathbf{s}, \mathbf{t}) = E(\mathcal{B}(\mathbf{s})\mathcal{B}(\mathbf{t})) = \prod_{j=1}^{d} \{s_j \wedge t_j\} - \prod_{i=1}^{d} \{s_j t_j\}$$

The eigenvalues and some eigenfunctions, corresponding to the kernel $\mathcal{K}_{\mathcal{X}}(\mathbf{s},\mathbf{t})$, can be derived with using of the eigenvalues and eigenfunctions, corresponding to the covariance function of the weighted multivariate Wiener process $\mathcal{W}(\mathbf{t})$

$$\mathcal{K}_{\mathcal{W}}(\mathbf{s},\mathbf{t}) = \prod_{j=1}^{d} \{s_j^{\beta_j} t_j^{\beta_j}\}\{s_j \wedge t_j\}.$$

In follows, we will use the notations

$$\nu_j = 1/(2(1+\beta_j)) > 0,\ j = 1, ..., d.$$

Karhunen-Loève expansion for $\mathcal{W}(\mathbf{t})$ is (see Deheuvels and Martynov, 2003).

$$\mathcal{W}(\mathbf{t}) = \sum_{k_1=1}^{\infty} \cdots \sum_{k_d=1}^{\infty} \sqrt{\lambda^*_{k_1 \ldots k_d}} Y_{k_1 \ldots k_d} e^*_{k_1 \ldots k_d}(\mathbf{t}), (4)$$

where

$$\lambda^*_{k_1 \ldots k_d} = \prod_{j=1}^{d} \left\{ 2\nu_j / z_{\nu_j - 1, k_j} \right\}^2$$

and

$$e^*_{k_1 \ldots k_d}(\mathbf{t}) = \prod_{j=1}^{d} \left[t_j^{\frac{1}{2\nu_j} - \frac{1}{2}} \left\{ \frac{J_{\nu_j}\left(z_{\nu_j - 1, k_j} t_j^{\frac{1}{2\nu_j}} \right)}{\sqrt{\nu_j} J_{\nu_j}\left(z_{\nu_j - 1, k_j} \right)} \right\} \right].$$

Here $J_\nu(\cdot)$ is the Bessel function, $0 < z_{\nu,1} < z_{\nu,2} < \ldots$ are the zeros of $J_\nu(\cdot)$ and $Y_{k_1 \ldots k_d}$: $k_1 \geq 1, \ldots, k_d \geq 1$ is an i.i.d. array of normal $N(0,1)$ random variables.

Now, we can extract from proof of the theorem in Krivjakova, Martynov and Tjurin (1977) and from Durbin (1970) the follow theorem:

The eigenvalues of $\mathcal{K}_{\mathcal{X}}(\mathbf{s},\mathbf{t})$ are

1. The λ_i^*, $i = 1, 2, ...$, with the multiplicity $s_i^* - 1$ for all λ_i^* with the multiplicity not equal to 1. 2. Maximal eigenvalue. 3. The solutions of the equation

$$\sum_{i=1}^{\infty} \frac{\sum_{k=1}^{s_i^*} (c_{i,k})^2}{\lambda_i^* - \lambda} = 1, \qquad (5)$$

where $c_{i,k}$ are some coefficients.

CRAMÉR–VON MISES TEST FOR THE GAUSSIAN PROCESS IN HILBERT SPACE

One of the problems for the goodness-of-fit tests is the problem to test if an observed random process $S(t)$ on $[0,1]$ is the Gaussian process with zero mean and a covariance function $K_S(t,\tau),\ t,\tau \in [0,1]$,

$$\int_0^1 K_S(t,t)dt < \infty,\ t,\tau \in [0,1]. \qquad (6)$$

The decision should be based on n observations $S_1(t), S_2(t), ..., S_n(t),\ t \in [0,1]$ of the random process $S(t)$.

Realizations of the process $S(t)$ are considered here as the elements of the separable Hilbert space $\mathcal{X}_{[0,1]} = L^2([0,1])$. As a basis for $\mathcal{X}_{[0,1]}$ we choose the orthonormal basis formed by eigenfunctions $g_1(t), g_2(t), ...$ of the integral operator

$$\int_0^1 K_S(t,\tau) g(\tau) d\tau. \tag{7}$$

Realization $S_i(t)$ of $S(t)$ can be represented as $(s_{i1}, s_{i2}, s_{i3}, ...) \in \mathcal{X}_{[0,1]}$, where

$$s_{ij} = \int_0^1 S_i(t) g_j(t) dt. \tag{8}$$

The processes $S(t)$ and $S_i(t)$ can be represented in the form of expansion in the mentioned basis as $\mathbf{s} = (s_1, s_2, s_3, ...)$.

The easiest way is to look at the problem in the general case. We will consider the probability space $(\mathcal{X},\ \mathcal{B},\ \nu)$ where $\mathcal{X}$ is a separable Hilbert space of elementary events, $\mathcal{B}$ is the σ-algebra of Borel set on $\mathcal{X}$ and ν is a probability measure. Let we have n observations $X^1, X^2, ..., X^n$ of the random element X of $(\mathcal{X},\ \mathcal{B},\ \nu)$. We will test hypothesis

$$H_0: \ \ \nu = \mu,$$

where μ is a Gaussian measure on $(\mathcal{X}, \mathcal{B})$ with a mean a and a covariance operator $K(z, w),\ z, w \in \mathcal{X}$. As a also $K(z, w)$ supposed be known. We can take $a = 0$.

Let $e = (e_1, e_2, ...)$ be the orthonormal basis of the eigenvectors of K and $\sigma_1^2, \sigma_2^2, ...$ be the eigenvalues of K. Let $x = (x_1, x_2, \ldots)$ be the representation of x in the basis e. Random element $X = (X_1, X_2, ...)$ has the independent components with the distributions $N(0, \sigma_i^2)$, $i = 1, 2, ...$ In result, we can transform the probability space $(\mathcal{X},\ \mathcal{B},\ \nu)$ to a probability space

$$([0,1]^\infty,\ C^\infty,\ \Gamma).$$

Here C is the Borel set on $[0,1]$ and Γ is the measure corresponding to ν.

Let Υ be the "uniform" measure on $([0,1]^\infty,\ C^\infty)$. Now we will test the hypothesis

$$H0:\ \Gamma = \Upsilon. \tag{9}$$

For application of the Cramér-von Mises-type test for testing the hypothesis (9), it need to introduce the function f on $[0,1]^\infty$ such, that it defines the measure Υ. In the finite dimensional case, as one variant of F can be chosen the obvious distribution function. In the considered case it is impossible. Instead, it can be proposed the function

$$F(t) = f_1(t_1)f_2(t_2)f_3(t_3)...,\quad t = (t_1, t_2, t_3, ...). \qquad (10)$$

This function must be nonzero for all points $t = (t_1, t_2, ...)$ in $(0,1]^\infty$, with of exception of a set of measure zero. The trasformed random variable X is defined now as $T = (T_1, T_2, ...)$.

The convenient example is

$$F(t) = P\{T_1 \le t_1^{r_1}, T_1 \le t_1^{r_1}, T_1 \le t_1^{r_1} ...\}$$
$$= t_1^{r_1} t_2^{r_2} t_3^{r_3} ..., \quad (11)$$

when r_i tends to zero sufficiently rapidly. Let $T^{(i)} = (T_{i1}, T_{i2}, ...)$ be the observations of T. The empirical function is

$$F_n(t) = \frac{1}{n} \sum_{i=1}^{\infty} 1_{T_{i1} \le t_1^{r_1},\, T_{i2}^{r_2}, ... \le t_2} \quad (12)$$

With this distribution function $F(t)$, $t \in [0,1]^\infty$, the measure μ can be restored.

The Cramér-von-Mises statistics is

$$\omega_n^2 = n \int_{[0,1]^\infty} \left(F_n(t) - \prod_{i=1}^{\infty} t_i^{r_i} \right)^2 dt_1 dt_2 ....,$$

The "empirical process"

$$\xi_n(t) = \sqrt{n}\left(F_n(t) - \prod_{i=1}^{\infty} t_i^{r_i}\right), \;\; t \in [0,1]^{\infty},$$

converges weakly in $L_2(\mathcal{X})$ to the Gaussian process with the covariance function

$$R(s,t) = \prod_{i=1}^{\infty} \min(s_i^{r_i}, t_i^{r_i}) - \prod_{i=1}^{\infty} s_i^{r_i} t_i^{r_i}, \; s,t \in [0,1]^{\infty}.$$

This assertion follows from the representation $\xi_n(t)$ by the sum

$$\xi_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\prod_{i=1}^{\infty} t_i^{r_i} - I_{T_i < t_i^{r_i}}\right), \;\; t \in [0,1]^{\infty}$$

of i.i.d. random functions and, for example, from Vaart and Wellner (1996)

If, for example,

$$r_i = i^{-v}, \ \ v > 1,$$

then the condition of the weak convergence

$$\int_{[0,1]^\infty} R(s,s)ds$$

$$= \prod_{i=1}^{\infty} \frac{1}{r_i + 1} - \prod_{i=1}^{\infty} \frac{1}{2r_i + 1} < \infty$$

is fulfilled. As $i \to \infty$ $r_i \downarrow 0$. It can be written

$$R(s,t) = R_0(s,t) - w(s)w(t),$$

where

$$R_0(s,t) = \prod_{i=1}^{\infty} \min(s_i^{r_i}, t_i^{r_i}),$$

$$w(s) = \prod_{i=1}^{\infty} s_i^{r_i}, \ \ s,t \in [0,1]^\infty.$$

Let μ_i and $\psi_i(\cdot)$, $i = 1, 2, \ldots$, the eigenvalues and eigenfunctions corresponding to $R_0(s,t)$. Then the eigenvalues $\lambda_1, \lambda_2, \ldots$ corresponding to $R(s,t)$ can be found from the equation

$$1 + \lambda \left(\sum_{i=1}^{\infty} \frac{q_i^2}{1 - \lambda/\mu_i} \right) = 0,$$

$$q_i = \int_{[0,1]^\infty} w(t)\psi_i(t)dt, \ i = 1, 2, \ldots .$$

Darling, D. A. (1955). The Cramér-Smirnov test in the parametric case. *Ann. Math. Statist.* **26** 1–20.

Now, we can consider the elementary kernel $r_i(t_i, s_i) = \min(t_i^{r_i}, s_i^{r_i})$, $i = 1, ...,$. His eigenfunctions and eigenvalues can be found from the integral equation

$$\phi(x) = \lambda \int_0^1 \min(x^{r_i}, y^{r_i})\phi(y)dy, x \in [0, 1].$$

By differentiation, we have

$$\phi'(t) = \lambda r_i t^{r_i - 1} \int_t^1 \phi(\tau)d\tau.$$

Let

$$z(t) = \int_t^1 \phi(\tau)d\tau.$$

The corresponding differential equation is

$$z''(t) + \lambda r_i t^{r_i - 1} z(t) = 0,$$

$$z(1) = z'(0) = 0.$$

The general solutions can be represented as the product of all possible combinations of eigenvalues and eigenfunctions.

The Cramér-von Mises statistic can be represented as

$$\omega_n^2 = n \int_{[0,1]^\infty} \left(\frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{\infty} I_{T_{i,j} < t_j^{r_i}} - \prod_{i=1}^{\infty} t_i^{r_i} \right) d\mathbf{t}.$$

The limit distribution of the Cramér-von Mises statistic can be calculated by the methods described above for multidimensional case. The statistic ω_n^2 can be calculated by the Monte-Carlo method. In turn, the distribution of the statistic was calculated also using the Monte Carlo method. We will present the estimated quantiles of the distribution ω_n^2 with $r_i = 1/i^3$. The integration was carried out over the cube $[0,1]^{10}$. The percent points are $P\{\omega_n^2 \leq 0.90\} \approx 0.34$ and $P\{\omega_n^2 \leq 0.95\} \approx 0.45$. It can be noted that corresponding percent points for the classical univariate Cramér-von Mises statistic are 0.35 and 0.46.

Deheuvels, P., Martynov, G. (2003) Karhunen-Loève expansions for weighted Wiener processes and Brownian bridges via Bessel functions., *Progress in Probability, Birkhäuser, Basel/Switzerland*, **55**, 57-93.

Martynov, G. V., (1979) The omega square tests, *Nauka, Moscow*, 80pp., (in Russian).

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

Krivjakova E. N., Martynov G. V., and Tjurin J. N. (1977) The distribution of the omega square statistics in the multivariate case. *"Theory of Probability and its Applications."*, **V. 22**, N 2, 415–420.

Martynov, G. V. (1975). Computation of distribution function of quadratic forms of normally distributed random variables. *Theor. Probab. Appl.* **20** 782–793.

Van der Vaart, A. W., Wellner, J. A. (1996) Weak Converge and Empirical Processes. Springer-Verlag, 508 pp.

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