Goodness-of-Fit Tests for Stochastic Processes with Parametric Basic Hypothesis

Yu.A. Kutoyants

Laboratoire de Statistique et Processus, Université du Maine

Le Mans, FRANCE

SAPS-IX, Le Mans, March 14, 2013

Classical GoF Tests

Simple Hypothesis

If we observe n i.i.d. r. v.'s $(X_1, \ldots, X_n) = X^n$ with distribution function $F(x)$ and the basic hypothesis is simple

$$\mathscr{H}_0, \qquad F(x) \equiv F_0(x), \qquad x \in \mathbb{R},$$

then the Cramér-von Mises statistic is

$$\Delta_n = n \int \left[\hat{F}_n(x) - F_0(x) \right]^2 \mathrm{d}F_0(x), \quad \hat{F}_n(x) = \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{\{X_j < x\}}$$

where $\hat{F}_n(x)$ is the empirical distribution function (EDF). Let us denote by $\mathcal{K}_\varepsilon$ the class of tests of asymptotic size ε, i.e.;

$$\mathcal{K}_\varepsilon = \left\{ \bar{\psi} \quad : \quad \mathbf{E}_0 \, \bar{\psi} = \varepsilon + o(1) \right\}.$$

We have the convergence (under hypothesis $\mathscr{H}_0$)

$$W_n(x) = \sqrt{n}\left(\hat{F}_n(x) - F_0(x)\right) \Longrightarrow W_0(F_0(x))$$

which provides

$$\Delta_n \Longrightarrow \Delta \equiv \int_0^1 W_0(s)^2 \, \mathrm{d}s,$$

where $W_0(\cdot)$ is Brownian bridge. Introduce the constant c_ε by the equation $\mathbf{P}\{\Delta > c_\varepsilon\} = \varepsilon$. Then the *Cramér-von Mises* test

$$\psi_n(X^n) = 1\!\!\mathrm{I}_{\{\Delta_n > c_\varepsilon\}} \in \mathcal{K}_\varepsilon, .$$

is

- *asymptotically distribution-free* (ADF), belongs to $\mathcal{K}_\varepsilon$,
- consistent against any alternative

$$\mathscr{H}_\rho = \{F(\cdot) : \|F(\cdot) - F_0(\cdot)\| \geq \rho\}, \qquad \rho > 0$$

Parametric Hypothesis

Suppose that the basic hypothesis is parametric:

$$\mathcal{H}_0 \quad : \qquad F(x) = F(\vartheta, x), \quad \vartheta \in \Theta$$

where $\Theta = (\alpha, \beta)$, i.e.; the d.f. $F(x)$ belongs to a parametric family $\mathcal{F} = \{F(\vartheta, x), \quad \vartheta \in \Theta\}$.

Introduce the statistic

$$\hat{\Delta}_n = n \int_{-\infty}^{\infty} \left[\hat{F}_n(x) - F\left(\hat{\vartheta}_n, x\right)\right]^2 \mathrm{d}F\left(\hat{\vartheta}_n, x\right),$$

where $\hat{\vartheta}_n$ is the maximum likelihood estimator (MLE). We have

$$\sqrt{n}\left(\hat{\vartheta}_n - \vartheta\right) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{\dot{l}(\vartheta, X_j)}{\mathrm{I}(\vartheta)} + o(1), \quad l(\vartheta, x) = \ln f(\vartheta, x).$$

We can write

$$
\begin{aligned}
U_n(x) &= \sqrt{n}\left(\hat{F}_n(x) - F\left(\hat{\vartheta}_n, x\right)\right) \\
&= \sqrt{n}\left(\hat{F}_n(x) - F(\vartheta, x)\right) + \sqrt{n}\left(F(\vartheta, x) - F\left(\hat{\vartheta}_n, x\right)\right) \\
&= W_n(x) - \sqrt{n}\left(\hat{\vartheta}_n - \vartheta\right)\, \dot{F}(\vartheta, x) + o(1) \Longrightarrow \\
& W_0(F(\vartheta, x)) - \int_{-\infty}^{\infty} \frac{\dot{l}(\vartheta, y)}{\sqrt{\mathrm{I}(\vartheta)}}\, \mathrm{d}W_0(F(\vartheta, y))\ \int_{-\infty}^{x} \frac{\dot{l}(\vartheta, y)}{\sqrt{\mathrm{I}(\vartheta)}}\, \mathrm{d}F(\vartheta, y) \\
&= W_0(s) - \int_0^1 h(v)\, \mathrm{d}W_0(v) \int_0^s h(v)\, \mathrm{d}v \equiv u(s),
\end{aligned}
$$

where

$$
s = F(\vartheta, x), \qquad h(s) = \frac{\dot{l}\left(\vartheta, F_\vartheta^{-1}(s)\right)}{\sqrt{\mathrm{I}(\vartheta)}}.
$$

Hence

$$\hat{\Delta}_n = \int U_n\left(x\right)^2 \mathrm{d}F\left(\hat{\vartheta}_n, x\right) \Longrightarrow \Delta\left(\vartheta, F\right) = \int_0^1 u\left(s\right)^2 \mathrm{d}s$$

where $u\left(s\right) = W_0\left(s\right) - \zeta H\left(s\right), \zeta \sim \mathcal{N}\left(0,1\right)$ and the C-vM type test $\hat{\psi}_n = 1\!\!\mathrm{I}_{\{\Delta_n > c_\varepsilon\}}$ The test is not ADF and the choice of the constant c_ε can be a difficult problem because $u\left(s\right) = u\left(s, \vartheta, F\right)$. We have to solve the equation

$$\mathbf{P}_\vartheta\left\{\int_0^1 u\left(s, \vartheta, F\right)^2 \mathrm{d}s > c_\varepsilon\left(\vartheta\right)\right\} = \varepsilon,$$

to show that $c_\varepsilon\left(\vartheta\right)$ is a continuous function and then to use $c_\varepsilon(\hat{\vartheta}_n)$.

There are at least two approaches allowing to avoid this problem. The first one is to find a transformation of U_n such that the limit statistics does not depend on ϑ (is *asymptotically parameter free* (APF) and the second to seek such transformation $L\left[U_n\right]$ that the test based on this statistic is ADF. The both approaches were applied.

Diffusion Processes (ergodic case)

Let $X^T = \{X_t, 0 \le t \le T\}$ be an observation of solution of some SDE

$$\mathrm{d}X_t = S\left(X_t\right)\,\mathrm{d}t + \sigma\left(X_t\right)\,\mathrm{d}W_t, \quad X_0, \quad 0 \le t \le T,$$

and we would like to know if this SDE is of the following form

$$\mathrm{d}X_t = S_0\left(X_t\right)\,\mathrm{d}t + \sigma\left(X_t\right)\,\mathrm{d}W_t, \quad X_0, \quad 0 \le t \le T,$$

where the trend $S_0\left(\cdot\right)$ and diffusion coefficient $\sigma\left(\cdot\right)^2$ are known functions. We suppose that this process is ergodic with invariant density $f_S\left(x\right)$.

Simple Basic Hypothesis.

We have to test the hypothesis

$$\mathscr{H}_0, \qquad S(x) = S_0(x).$$

Our goal is to construct the GoF tests, which have the properties:

- *belong to $\mathcal{K}_\varepsilon$,*
- *are asyptotically distribution free,*
- *are consistent against a wide classe of alternatives*

The tests are based on EDF and on empirical density function (ED)

$$\hat{F}_T(x) = \frac{1}{T}\int_0^T 1\!\!\mathrm{I}_{\{X_t<x\}}\,\mathrm{d}t, \qquad \hat{f}_T(x) = \frac{\Lambda_T(x)}{T\sigma(x)^2}$$

The tests studied are direct analogues of the classical Cramér-von Mises test. Introduce the statistics

$$\begin{aligned}\Delta_T &= T\int_{-\infty}^{\infty}\left[\hat{F}_T(x)-F_{S_0}(x)\right]^2\,\mathrm{d}F_{S_0}(x)\\ &= \int_0^T\left[\hat{F}_T(X_t)-F_{S_0}(X_t)\right]^2\mathrm{d}t,\\ \delta_T &= T\int_{-\infty}^{\infty}\left[\hat{f}_T(x)-f_{S_0}(x)\right]^2\,\mathrm{d}F_{S_0}(x)\\ &= \int_0^T\left[\hat{f}_T(X_t)-f_{S_0}(X_t)\right]^2\mathrm{d}t+o(1)\,.\end{aligned}$$

The tests are

$$\Psi\left(X^T\right)=1\!\!1_{\{\Delta_T>c_\varepsilon\}},\qquad \psi\left(X^T\right)=1\!\!1_{\{\delta_T>d_\varepsilon\}},$$

where $\lim_{T\to\infty}\mathbf{E}_{S_0}\Psi\left(X^T\right)=\varepsilon$, $\lim_{T\to\infty}\mathbf{E}_{S_0}\psi\left(X^T\right)=\varepsilon$.

Problem: how to find $c_\varepsilon, d_\varepsilon$?

The random functions $\eta_T(x) = \sqrt{T}\left(\hat{F}_T(x) - F_S(x)\right)$, $\zeta_T(x) = \sqrt{T}\left(\hat{f}_T(x) - f_S(x)\right)$ admit the representations

$$\eta_T(x) = \frac{2}{\sqrt{T}} \int_0^T \frac{F_S(X_t)\, F_S(x) - F_S(X_t \wedge x)}{\sigma(X_t)\, f_S(X_t)}\, \mathrm{d}W_t + o(1),$$

$$\zeta_T(x) = \frac{2 f_S(x)}{\sqrt{T}} \int_0^T \frac{F_S(X_t) - 1\!\!1_{\{X_t > x\}}}{\sigma(X_t)\, f_S(X_t)}\, \mathrm{d}W_t + o(1)$$

and have the limits

$$\eta_T(x) \Longrightarrow \eta(x) = 2 \int_{-\infty}^{\infty} \frac{F_S(y)\, F_S(x) - F_S(y \wedge x)}{\sigma(y) \sqrt{f_S(y)}}\, \mathrm{d}W(y),$$

$$\zeta_T(x) \Longrightarrow \zeta(x) = 2 f_S(x) \int_{-\infty}^{\infty} \frac{F_S(y) - 1\!\!1_{\{y > x\}}}{\sigma(y) \sqrt{f_S(y)}}\, \mathrm{d}W(y),$$

The tests are obviously not ADF.

Our goal is to find a linear transformation $L(\zeta_T)$ of the random function $\zeta_T(x)$ such that

$$\delta_T^*\left(X^T\right) = \int \left[L\left[\zeta_T\right](x)\right]^2 \mathrm{d}F_{S_0}(x) \Longrightarrow \delta^* \equiv \int_0^1 w(s)^2\, \mathrm{d}s,$$

where $w(s), 0 \leq s \leq 1$ is a Wiener process. Then obviously the test $\psi_T^*\left(X^T\right) = 1\!\!1_{\left\{\delta_T^*(X^T)>c_\varepsilon\right\}}$ with c_ε from the equation $\mathbf{P}\left\{\delta^* > c_\varepsilon\right\} = \varepsilon$ will be ADF.

We study the test based on $\zeta_T(\cdot)$ only because $\zeta_T(\cdot) = \eta_T'(\cdot)$ (linear operation).

We can rewrite the stochastic integral as follows

$$\int_{-\infty}^{\infty} \frac{F_{S_0}(y) - \mathbb{I}_{\{y>x\}}}{\sigma(y)\sqrt{f_{S_0}(y)}} \, \mathrm{d}W(y)$$

$$= \int_0^1 \frac{s - \mathbb{I}_{\{s>t\}}}{a(s)\, b(s)} \, \mathrm{d}w(s)$$

$$= \int_0^t \frac{s}{a(s)\, b(s)} \, \mathrm{d}w(s) + \int_t^1 \frac{s-1}{a(s)\, b(s)} \, \mathrm{d}w(s) \equiv u(t),$$

where $F_{S_0}(y) = s, \qquad F_{S_0}(x) = t, \quad a(s) = \sigma\left(F_{S_0}^{-1}(s)\right),$

$$w(s) = \int_{-\infty}^{F_{S_0}^{-1}(s)} \sqrt{f_{S_0}(y)}\, \mathrm{d}W(y), \quad b(s) = f_{S_0}\left(F_{S_0}^{-1}(s)\right).$$

Here $F_{S_0}^{-1}(s)$ is the function inverse to $F_{S_0}(y)$, i.e., the solution y of the equation $F_{S_0}(y) = s$.

The random function $u(\cdot)$ has the differential

$$\mathrm{d}u(t) = \frac{t}{a(t)\,b(t)}\,\mathrm{d}w(t) - \frac{t-1}{a(t)\,b(t)}\,\mathrm{d}w(t) = \frac{1}{a(t)\,b(t)}\,\mathrm{d}w(t)\,.$$

Hence the integral (understood in the mean square sense)

$$\int_0^t a(s)\,b(s)\;\mathrm{d}u(s) = w(t)$$

provides us the desired transformation. Indeed, we have

$$u(t) = \frac{\zeta\left(S_0, F_{S_0}^{-1}(t)\right)}{2f_{S_0}\left(F_{S_0}^{-1}(t)\right)}$$

and

$$\int_0^t a(s)\,b(s)\;\mathrm{d}u(s) = \int_{-\infty}^x \sigma(y)\,f_{S_0}(y)\;\mathrm{d}\left[\frac{\zeta(S_0, y)}{2f_{S_0}(y)}\right] = w\left(F_{S_0}(x)\right).$$

Therefore, we can write

$$\delta^* = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{x} \sigma(y)\, f_{S_0}(y)\ \mathrm{d}\left[\frac{\zeta(S_0, y)}{2 f_{S_0}(y)} \right] \right)^2 \mathrm{d}F_{S_0}(x) = \int_0^1 w(t)^2\ \mathrm{d}t.$$

Let us introduce the statistic

$$\delta_T^*\left(X^T\right) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{x} \sigma(y)\, f_{S_0}(y)\ \mathrm{d}\left[\frac{\zeta_T(S_0, y)}{2 f_{S_0}(y)} \right] \right)^2 \mathrm{d}F_{S_0}(x),$$

where we have to define the integral with respect to the normalized empirical density

$$\int_a^b h(x)\ \mathrm{d}\left[\frac{\zeta_T(x)}{2 f_{S_0}(x)} \right].$$

Using the representation for $\zeta_T(x_{i+1})$ for any piecewise continuous function $h(x)$ with bounded support and partition $a = x_0 < x_1 < \ldots < x_m = b$ we can write

$$\sum_{x_i} h(\tilde{x}_i)\left[\frac{\zeta_T(x_{i+1})}{2f_{S_0}(x_{i+1})} - \frac{\zeta_T(x_i)}{2f_{S_0}(x_i)}\right]$$

$$= \frac{1}{\sqrt{T}}\int_0^T \frac{\sum_{x_i} h(\tilde{x}_i)\,\mathbb{I}_{\{x_i < X_t \le x_{i+1}\}}}{\sigma(X_t)\, f_{S_0}(X_t)}\,\mathrm{d}W_t$$

$$- \frac{1}{\sqrt{T}}\int_{X_0}^{X_T} \frac{\sum_{x_i} h(\tilde{x}_i)\,\mathbb{I}_{\{x_i < y \le x_{i+1}\}}}{\sigma(y)\, f_{S_0}(y)}\,\mathrm{d}y.$$

Therefore as $\max |x_{i+1} - x_i| \to 0$ we obtain the limit

$$\lim_{m \to \infty} \sum_{x_i} h(\tilde{x}_i) \left[\frac{\zeta_T(x_{i+1})}{2 f_{S_0}(x_{i+1})} - \frac{\zeta_T(x_i)}{2 f_{S_0}(x_i)} \right]$$

$$= \frac{1}{\sqrt{T}} \int_0^T \frac{h(X_t) \mathbb{I}_{\{a < X_t \leq b\}}}{\sigma(X_t) f_{S_0}(X_t)} \, \mathrm{d}W_t - \frac{1}{\sqrt{T}} \int_{X_0}^{X_T} \frac{h(y) \mathbb{I}_{\{a < y \leq b\}}}{\sigma(y)^2 f_{S_0}(y)} \, \mathrm{d}y.$$

If $a = -\infty$ (our case) then this limit exists for a class of functions vanishing in $-\infty$ because $\sigma(y) f_{S_0}(y) \to 0$ as $|y| \to \infty$.

Remind that in our case $h(x) = \sigma(x) f_{S_0}(x)$ and the integral

$$\int_{-\infty}^{x} \sigma(y) f_{S_0}(y) \; \mathrm{d}\left[\frac{\zeta_T(S_0, y)}{2 f_{S_0}(y)}\right]$$
$$= \frac{1}{\sqrt{T}} \int_0^T \mathbb{1}_{\{X_t \le x\}} \, \mathrm{d}W_t - \frac{1}{\sqrt{T}} \int_{X_0}^{X_T} \frac{\mathbb{1}_{\{y \le x\}}}{\sigma(y)} \, \mathrm{d}y.$$

Therefore,

$$\delta_T(X_T) = \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{T}} \int_0^T \mathbb{1}_{\{X_t \le x\}} \, \mathrm{d}W_t - \frac{H(x, X_0, X_T)}{\sqrt{T}} \right)^2 \mathrm{d}F_{S_0}(x)$$
$$= \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{T}} \int_0^T \mathbb{1}_{\{X_t \le x\}} \, \mathrm{d}W_t \right)^2 \mathrm{d}F_{S_0}(x) + O\left(\frac{1}{\sqrt{T}}\right),$$

where we put

$$H(x, y, z) = \int_y^z \frac{\mathbb{1}_{\{v \le x\}}}{\sigma(v)} \, \mathrm{d}v.$$

When we know the form of the statistic δ_T we can construct another goodness of fit test with the same asymptotics as follows. Let us introduce the statistic

$$\delta_T^* = \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{T}} \int_0^T \frac{\mathbb{1}_{\{X_t < x\}}}{\sigma(X_t)} \left[\mathrm{d}X_t - S_0(X_t)\,\mathrm{d}t \right] \right)^2 \mathrm{d}F_{S_0}(x)$$

and the constant c_ε:

$$\mathbf{P}\left\{ \int_0^1 w(t)^2 \,\mathrm{d}t > c_\varepsilon \right\}.$$

Introduce the test

$$\hat{\psi}_T = \mathbb{1}_{\{\delta_T^* > c_\varepsilon\}}$$

The similar statistic was used by Negri *et al.* (2009) for ADF K-S type test.

We suppose that under (nonparametric) alternative

$$\mathcal{H}_1 \qquad : \qquad \mathbf{E}_S \left(\frac{S(\xi) - S_0(\xi)}{\sigma(\xi)} \right)^2 > 0$$

the function $S(\cdot)$ satisfies the usual conditions Here ξ is the random variable with the density $f_S(x)$.

Proposition 1 (K. 2012) *The test* $\hat{\psi}_T \in \mathcal{K}_\varepsilon$*, is ADF and is consistent against alternative* $\mathcal{H}_1$*.*

Composite Basic Hypothesis

We observe an ergodic diffusion process $X^T = (X_t, 0 \leq t \leq T)$ solution of the equation

$$\mathrm{d}X_t = S(X_t)\ \mathrm{d}t + \sigma(X_t)\ \mathrm{d}W_t, \qquad X_0, \quad 0 \leq t \leq T$$

and we have to test the composite basic hypothesis $\mathcal{H}_0$, that this process admits the stochastic differential

$$\mathrm{d}X_t = S(\vartheta, X_t)\ \mathrm{d}t + \sigma(X_t)\ \mathrm{d}W_t, \qquad X_0, \quad 0 \leq t \leq T,$$

where $\vartheta \in \Theta \subset R^k$ is some (unknown) parameter.

Our goal is to construct such tests $\hat{\psi}_T = 1\!\!1_{\{\Delta_T > c_\varepsilon\}}$ that

$$\lim_{T\to\infty} \mathbf{E}_\vartheta \hat{\psi}_T = \varepsilon.$$

for all $\vartheta \in \Theta$. We consider two approaches: APF and ADF.

Asymptotically Parameter Free Tests.

Problem: how to choose such Δ_T that its limit in distribution Δ does not depend on ϑ? We call the corresponding test *asymptotically parameter free* (APF). Let us introduce a parametric family of ergodic diffusion processes (basic composite hypothesis $\mathcal{H}_0$)

$$\mathrm{d}X_t = -\beta\ \mathrm{sgn}\left(X_t - \alpha\right)\left|X_t - \alpha\right|^{\gamma}\ \mathrm{d}t + \sigma\,\mathrm{d}W_t, \qquad X_0, \quad 0 \leq t \leq T,$$

where $\vartheta = (\alpha, \beta)$ is the unknown parameter,
$\vartheta \in \Theta = (a_1, a_2) \times (b_1, b_2)$ and we suppose that $b_1 > 0$. The alternative (nonparametric) is
$\mathcal{H}_1$: *the observed process has a trend coefficient* $S\left(x\right)$ *which does not belong to this parametric family* $\left\{S\left(\vartheta, x\right), \vartheta \in \Theta\right\}$.

Note that if $\gamma = 1$, then we obtain *Ornstein-Uhlenbeck* process

$$\mathrm{d}X_t = -\beta \ (X_t - \alpha) \ \mathrm{d}t + \sigma \, \mathrm{d}W_t, \qquad X_0, \quad 0 \le t \le T$$

and if $\gamma = 0$ then the solution is *simple switching process*

$$\mathrm{d}X_t = -\beta \ \mathrm{sgn}\,(X_t - \alpha) \ \mathrm{d}t + \sigma \, \mathrm{d}W_t, \qquad X_0, \quad 0 \le t \le T.$$

If $\gamma = 3$, then we obtain another well-known process

$$\mathrm{d}X_t = -\beta \ (X_t - \alpha)^3 \ \mathrm{d}t + \sigma \, \mathrm{d}W_t, \qquad X_0, \quad 0 \le t \le T.$$

The invariant density is

$$f(\vartheta, x) = \frac{\beta^{\frac{1}{\gamma+1}}}{G_\gamma\ \sigma^{\frac{2}{\gamma+1}}}\ \exp\left\{-\frac{2\beta\left|x-\alpha\right|^{\gamma+1}}{(\gamma+1)\,\sigma^2}\right\}.$$

Its form suggests the change of variables

$$z = \frac{\beta^{\frac{1}{\gamma+1}}}{\sigma^{\frac{2}{\gamma+1}}}\,(x-\alpha)\,, \qquad Z_t = \frac{\beta^{\frac{1}{\gamma+1}}}{\sigma^{\frac{2}{\gamma+1}}}\,(X_t-\alpha)\,.$$

The process Z_t has the stoch. differential

$$\mathrm{d}Z_t = -\,\mathrm{sgn}\,(Z_t)\ \left|Z_t\right|^\gamma\ \mathrm{d}\left(t\beta^{\frac{2}{\gamma+1}}\sigma^{\frac{2(\gamma-1)}{\gamma+1}}\right)+\beta^{\frac{1}{\gamma+1}}\sigma^{\frac{\gamma-1}{\gamma+1}}\mathrm{d}W_t, \quad 0\le t\le T.$$

Therefore, if we denote

$$Y_s = Z_{s\beta^{-\frac{2}{\gamma+1}}\sigma^{-\frac{2(\gamma-1)}{\gamma+1}}}, \qquad 0 \leq s = t\beta^{\frac{2}{\gamma+1}}\sigma^{\frac{2(\gamma-1)}{\gamma+1}} \leq T_* = T\beta^{\frac{2}{\gamma+1}}\sigma^{\frac{2(\gamma-1)}{\gamma+1}},$$

the process Y_s satisfies the equation

$$\mathrm{d}Y_s = -\operatorname{sgn}(Y_s)\;|Y_s|^{\gamma}\,\mathrm{d}s + \mathrm{d}w_s, \qquad Y_0, \quad 0 \leq s \leq T_*,$$

where $w_s = \beta^{\frac{1}{\gamma+1}}\sigma^{\frac{\gamma-1}{\gamma+1}}W_t$ (here $t = s\beta^{-\frac{2}{\gamma+1}}\sigma^{-\frac{2(\gamma-1)}{\gamma+1}}$) is another Wiener process. Obviously the process Y_s is ergodic with the invariant density $f_0(x)$.

This process does not depend on ϑ and has invariant density $f_0(z)$ which corresponds $\alpha = 0$, $\beta = 1$ and $\sigma = 1$.

We have

$$\hat{\eta}_T(x) = \sqrt{T}\left(\hat{F}_T(x) - F\left(\hat{\vartheta}_T, x\right)\right)$$
$$= \sqrt{T}\left(\hat{F}_T(x) - F(\vartheta, x)\right) + \sqrt{T}\left(F(\vartheta, x) - F\left(\hat{\vartheta}_T, x\right)\right)$$
$$= \eta_T(x) - \sqrt{T}\left(\hat{\alpha}_T - \alpha\right)\frac{\partial F(\vartheta, x)}{\partial \alpha} - \sqrt{T}\left(\hat{\beta}_T - \beta\right)\frac{\partial F(\vartheta, x)}{\partial \beta} + o(1).$$

Now we express all terms of this representation as functions of Z_t and we change the variables $x \to z$. Then we see that all three parts have the same factors

$$\beta^{\frac{1}{\gamma+1}}\sigma^{\frac{(\gamma-1)}{\gamma+1}}.$$

Case $\gamma \geq 1$. Introduce the first test based on EDF and MLE:

$$\hat{\psi}_T = \mathbb{I}_{\{\Delta_T(X^T) > c_\varepsilon\}}$$

, where

$$\Delta_T\left(X^T\right) = \hat{\beta}_T^{\frac{2}{\gamma+1}} \sigma^{\frac{2(\gamma-1)}{\gamma+1}} T \int_{-\infty}^{\infty} \left[\hat{F}_T(x) - F\left(\hat{\vartheta}_T, x\right)\right]^2 \mathrm{d}F\left(\hat{\vartheta}_T, x\right).$$

We have

$$\Delta_T\left(X^T\right) \Longrightarrow \Delta = \int_{-\infty}^{\infty} \left[\Phi(y) + y\, f_0(y)\ \Psi - f_0(y)\ \Pi\right]^2 f_0(y)\,\mathrm{d}y,$$

Hence the test $\hat{\psi}_T$ with c_ε from the equation $\mathbf{P}(\Delta > c_\varepsilon) = \varepsilon$ is APF.

Here $\Phi(y)$, Ψ and Π do not depend on ϑ and σ.

$$\Phi(y) = 2\int_{-\infty}^{\infty} \frac{F_0(z)\,F_0(y) - F_0(z \wedge y)}{\sqrt{f_0(y)}}\ \mathrm{d}W(z),$$

$$\Psi = \frac{1}{(\gamma+1)\,\mathbf{E}_0\,|\xi|^{2\gamma}} \int_{-\infty}^{\infty} \operatorname{sgn}(z)\,|z|^{\gamma}\,f_0(z)\ \mathrm{d}W(z),$$

$$\Pi = \frac{1}{\gamma\mathbf{E}_0\,|\xi|^{2\gamma-2}} \int_{-\infty}^{\infty} \operatorname{sgn}(z)\,|z|^{\gamma-1}\,f_0(z)\ \mathrm{d}W(z).$$

The r.v. ξ has the density function $f_0(z)$ and $W(\cdot)$ is two-sided Wiener process.

Let us consider the test based on empirical density

$$\tilde{\psi}_T = 1\!\!\mathrm{I}_{\{\delta_T(X^T)>c_\varepsilon\}},$$

where the test statistic is

$$\delta_T\left(X^T\right) = \sigma^2 T \int_{-\infty}^{\infty} \left[\hat{f}_T\left(x\right) - f\left(\hat{\vartheta}_T, x\right)\right]^2 \mathrm{d}F\left(\hat{\vartheta}_T, x\right).$$

Then once more we have limit $\delta_T\left(X^T\right) \Longrightarrow \delta$ where

$$\delta = \int_{-\infty}^{\infty} \left[\tilde{\Phi}\left(y\right) - 2\,\mathrm{sgn}\left(y\right)\left|y\right|^{\gamma}\Pi - \left[1 - 2\left|y\right|^{\gamma+1}\right]\Psi\right]^2 f_0\left(y\right)^3 \mathrm{d}y,$$

Hence the test $\tilde{\psi}_T$ with corresponding c_ε is APF.

Case $0 \leq \gamma < \frac{1}{2}$.

The difference is in the behavior of the MLE of the component α. Now $\hat{\alpha}_T$ converges to α with the rate better than $\sqrt{T}$:

$$T^{\frac{1}{2\gamma+1}} \left(\hat{\alpha}_T - \alpha\right) \Longrightarrow \hat{u} = \arg\sup Z\left(u\right),$$

where for $0 < \gamma < 1/2$ we have

$$Z\left(u\right) = \exp\left\{W^H\left(u\Gamma_\vartheta\right) - \frac{\left|u\Gamma_\vartheta\right|^H}{2}\right\}$$

Here $W^H\left(\cdot\right)$ is two-sided fBm. The limit distributions for test statistics we obtain if we put $\Pi = 0$. Another $Z\left(\cdot\right)$ but the same limits we have if $\gamma = 0$.

Asymptotically Distribution Free Tests.

(joint work with M.Kleptsyna (Le Mans) and R. Liptser (Tel Aviv))

Remind first what happens in i.i.d. case. Suppose that the basic hypothesis is

$$\mathcal{H}_0 \qquad\qquad F(x) = F_0(\vartheta, x)\,,, \quad \vartheta \in \Theta.$$

The test statistic is

$$\hat{\Delta}_n = n \int \left[\hat{F}_n(x) - F_0\left(\hat{\vartheta}_n, x\right)\right]^2 \mathrm{d}F_0\left(\hat{\vartheta}_n, x\right),$$

where $\hat{\vartheta}_n$ is the MLE. We know that

$$\hat{\Delta}_n \Longrightarrow \int_0^1 \left[W^0(t) - \zeta\, H(t)\right]^2 \mathrm{d}t$$

Problem: how to find a linear transformation $L(U)(t)$ of the process $U(t) = W^0(t) - \zeta\, H(t)$ such that for

$$\eta_n(x) = \sqrt{n}\left(\hat{F}_n(x) - F_0\left(\hat{\vartheta}_n, x\right)\right)$$

we have

$$\tilde{W}_n^2 = \int \left[L(\eta_n)(x)\right]^2 \mathrm{d}F_0\left(\hat{\vartheta}_n, x\right) \Longrightarrow \int_0^1 \left[L(U)(t)\right]^2 \mathrm{d}t = \int_0^1 w(t)^2\, \mathrm{d}t.$$

Such transformation was proposed by Khmaladze in 1980 and is based on two strong results. One of Shepp (1966) (equivalence of Gaussian measures) and the second of Hitsuda (1968) (representation of Gaussian processes equivalent to Wiener process). First gives the condition for the second and together yield this linear representation.

For ergodic diffusion processes we have the parametric family (under hypothesis $\mathcal{H}_0$)

$$\mathrm{d}X_t = S\left(\vartheta, x\right)\mathrm{d}t + \sigma\left(X_t\right)\mathrm{d}W_t, \quad X_0 \quad 0 \leq t \leq T$$

where $\vartheta \in \Theta \subset R^d$ and use the test statistics based on empirical density

$$\hat{\zeta}_T\left(x\right) = \sqrt{T}\left(\hat{f}_T\left(x\right) - f\left(\hat{\vartheta}_T, x\right)\right).$$

We have the similar expansion

$$\begin{aligned}
\hat{\zeta}_T\left(x\right) &= \sqrt{T}\left(\hat{f}_T\left(x\right) - f\left(\vartheta, x\right)\right) + \sqrt{T}\left(f\left(\vartheta, x\right) - f\left(\hat{\vartheta}_T, x\right)\right) \\
&= \zeta_T\left(x\right) - \sqrt{T}\left(\hat{\vartheta}_T - \vartheta\right)\, \dot{f}\left(\vartheta, x\right) + o\left(1\right) \\
&\Longrightarrow \hat{\zeta}\left(x\right) = \zeta\left(x\right) - \zeta \mathrm{I}\left(\vartheta\right)^{-1/2}\, \dot{f}\left(\vartheta, x\right) \\
&\qquad = u\left(t\right) - \zeta g\left(t\right) = \hat{u}\left(t\right).
\end{aligned}$$

Therefore our goal is to find a linear transformation $L(\hat{u})(t) = w(t)$. Then

$$\delta_T = \int_{-\infty}^{\infty} \left[L\left(\hat{\zeta}_T\right)(x)\right]^2 \mathrm{d}F\left(\hat{\vartheta}_T, x\right)$$

$$\Longrightarrow \int_0^1 \left[L(\hat{u})(t)\right]^2 \mathrm{d}t = \int_0^1 w(t)^2 \,\mathrm{d}t$$

and the test

$$\hat{\psi}_T = 1\!\!1_{\{\delta_T > c_\varepsilon\}} \in \mathcal{K}_\varepsilon \qquad \mathbf{P}\left\{\int_0^1 w(t)^2 \,\mathrm{d}t > c_\varepsilon\right\} = \varepsilon.$$

First we apply the linear transformation $L = L_1$ as above

$$L[\hat{u}](t) = w(t) - \zeta H(t) = \hat{v}(t).$$

Then we reduce this problem to the solution of Fredholm equation with degenerate kernel and obtain the second transformation

$$L_2\left[\hat{v}\right](t) = \hat{v}(t) + \int_0^t \frac{h(s)}{N(s)} \int_0^s h(r)\, \mathrm{d}\hat{v}(r)\ \mathrm{d}s = w(t)$$

Here

$$N(s) = \int_0^s h(r)^2\, \mathrm{d}r.$$

Therefore if we define $L = L_1 * L_2$ then

$$\hat{\delta}_T = \int_{-\infty}^{\infty} \left[L\left(\hat{\zeta}_T\right)(x)\right]^2 \mathrm{d}F\left(\hat{\vartheta}_T, x\right) \Longrightarrow \int_0^1 w(t)^2\, \mathrm{d}t$$

and the test

$$\hat{\psi}_T = 1\!\!1_{\left\{\hat{\delta}_T > c_\varepsilon\right\}} \in \mathcal{K}_\varepsilon$$

Small noise asymptotics.

Model

$$\mathrm{d}X_t = S\left(\vartheta, X_t\right)\mathrm{d}t + \varepsilon\sigma\left(X_t\right)\mathrm{d}W_t, \qquad X_0 = x_0, 0 \le t \le T$$

where $\varepsilon \rightarrow$. The limit is

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = S\left(\vartheta, x_t\right), \qquad x_0, 0 \le t \le T$$

We have $x_t = x_t\left(\vartheta\right)$ and the statistic is

$$\Delta_\varepsilon = \varepsilon^{-2}\int_0^T \left[X_t - x_t\left(\hat\vartheta_\varepsilon, r\right)\right]^2 \mathrm{d}t$$

It converges to the process

$$\Delta_\varepsilon \Longrightarrow \int_0^T \left[x_t^{(1)} - \zeta\dot x_t\left(\vartheta\right)\right]^2 \mathrm{d}t$$

We propose the linear transformation wich makes the GoF test ADF.

References

[1] Dachian, S. and Kutoyants, Yu.A. (2007) On the goodness-of-fit tests for some continuous time processes, in *Statistical Models and Methods for Biomedical and Technical Systems*, F.Vonta *et al.* (Eds), Birkhäuser, Boston, 395-413.

[2] Kleptsyna, M., Kutoyants Yu. A., Liptser R. (2012) On asymptotically distribution free tests with parametric hypothesis for ergodic diffusion processes.

[3] Kutoyants, Yu.A. (2004) Statistical Inference for Ergodic Diffusion Processes, Springer, London.

[4] Kutoyants Yu. A., (2010) On the goodness-of-fit testing for ergodic diffusion processes. Journal of Nonparametric Statistics, 22, 4, 529-543.

[5] Kutoyants, Yu. A., (2011) On goodness-of-fit tests for perturbed dynamical systems. Journal of Statistical Planning and Inference, 141,1655-1666.

[6] Kutoyants, Yu. A., (2012) On asymptotically distribution and parameter free tests for ergodic diffusion processes, Submitted.