Pseudo-Divergence Test Statistics for Multidimensional Diffusion Processes Observed at Discrete Times

S.M. Iacus & A. De Gregorio
Universities of Milan & Rome

SAPS VII, 19 March 2009, Le Mans

Divergences are measure of discrepancy between statistical models.

Let $\{P_\theta : \theta \in \Theta\}$ be a family of probability measures on a measurable space $(\mathcal{X}, \mathcal{A})$, $\theta \subset R^d$, $d \geq 1$.

The measures P_θ are assumed to have densities p_θ w.r.t some common dominating measure μ on $\mathcal{X}$

$$p(x, \theta) = \frac{\mathrm{d}P_\theta}{\mathrm{d}\mu}(x)$$

ϕ-divergences are defined as follows

$$D_\phi(\theta, \theta_0) = \int_{\mathcal{X}} p(\theta_0, x)\phi\left(\frac{p(\theta, x)}{p(\theta_0, x)}\right)\mu(\mathrm{d}x)$$
$$= \mathbf{E}_{\theta_0}\phi\left(\frac{p(X, \theta)}{p(X, \theta_0)}\right)$$

and the minimal requirement on the function ϕ is: $\phi(1) = 0$

The ϕ-divergences where introduced by Csiszár (1963) and studied extensively later in Liese and Vajda (1987)

They include most of known other divergences. We discuss some examples in the following

The α-divergences (Csiszár, 1967, Amari, 1985) are defined as

$$D_\alpha(\theta, \theta_0) = D_{\phi_\alpha}(\theta, \theta_0)$$

with

$$\phi_\alpha(x) = \frac{4\left(1 - x^{\frac{1+\alpha}{2}}\right)}{1-\alpha^2}, \quad -1 < \alpha < 1$$

They are such that $D_\alpha(\theta_0, \theta) = D_{-\alpha}(\theta, \theta_0)$.

The α-divergences include some special cases.

For example, for $\alpha \to -1$, D_{-1} is the well-known Kullback-Leibler divergence

$$D_{-1}(\theta, \theta_0) = D_{KL}(\theta, \theta_0) = E_{\theta_0} \log\left(\frac{p(X, \theta_0)}{p(X, \theta)}\right)$$

For $\alpha \to 0$, the Hellinger distance (see, e.g., Beran, 1977, Simpson, 1989) can be derived

$$d_H(\theta, \theta_0) = \frac{1}{2} E\left(\sqrt{p(X, \theta)} - \sqrt{p(X, \theta_0)}\right)^2$$

Examples: Rényi divergences

The α-divergence is also equivalent to the Rényi's divergence (Rényi, 1961)

$$R_\alpha(\theta, \theta_0) = \frac{1}{1-\alpha} \log E_{\theta_0} \left(\frac{p(X, \theta)}{p(X, \theta_0)} \right)^\alpha$$

$$D_{KL}(\theta, \theta_0) = \lim_{\alpha \to 1} R_\alpha(\theta, \theta_0)$$

again

$$d_H(\theta, \theta_0) = 1 - \exp\left\{ \frac{1}{2} R_{\frac{1}{2}}(\theta, \theta_0) \right\}$$

Liese and Vajda (1987) generalized Rényi divergences to all real orders $\alpha \neq 0, 1$

$$D_\alpha(\theta, \theta_0) = \frac{1}{\alpha(\alpha - 1)} \log E_{\theta_0} \left(\frac{p(X, \theta)}{p(X, \theta_0)} \right)^\alpha$$

only for $\alpha = \frac{1}{2}$ the divergence is symmetric

$$D_{\frac{1}{2}}(\theta_0, \theta) = D_{\frac{1}{2}}(\theta, \theta_0) = 4 \log \int \sqrt{p(x, \theta) p(\theta_0, x)} \mu(\mathrm{d}x)$$

[known as Bhattacharyya (1946) divergence], otherwise

$$D_\alpha(\theta_0, \theta) = D_{1-\alpha}(\theta, \theta_0)$$

The transformation

$$\psi(R_\alpha) = (\exp\{(\alpha-1)R_\alpha - 1\}/(1-\alpha)$$

coincides with the power-divergence introduced by Cressie and Read (1984)

Power divergences D_{ϕ_λ} can be obtained directly from the ϕ-divergences choosing

$$\phi_\lambda(x) = \frac{x^{\lambda+1} - \lambda(x-1) - x}{\lambda(\lambda+1)}, \quad \lambda \in \mathbb{R} - \{0, -1\}$$

...and so forth

Summary of ϕ-divergences (see Pardo, 2006)

$\phi(x)$ with $x = p(\theta, \cdot)/p(\theta_0, \cdot)$	*Divergence*
$x \log x - x + 1$	Kullback-Leibler
$- \log x + -1$	Minimum Discrimination Information
$(x-1) \log x$	J-divergence
$\frac{1}{2}(x-1)^2$	Pearson, Kagan
$\frac{(x-1)^2}{(x+1)^2}$	Balakrishnan & Sanghvi
$\frac{-x^s + s(x-1)+1}{1-s}, \quad s \neq 1$	Rathie & Kannappan
$\frac{1-x}{2} - \left(\frac{1+x^{-r}}{2}\right)^{-1/r} \quad r > 0$	Harmonic mean (Mathai & Rathie)
$\frac{(1-x)^2}{2(a+(1-a)x)} \quad 0 \leq a \leq 1$	Rukhin
$\frac{ax \log x - (ax+1-a) \log(ax+1-a)}{a(1-a)} \quad a \neq 0, 1$	Lin
$\frac{x^{\lambda+1} - x - \lambda(x-1)}{\lambda(\lambda+1)} \quad \lambda \neq 0, -1$	Cressie & Read
$\lvert 1 - x^a \rvert^{1/a} \quad 0 < a < 1$	Matusita
$\lvert 1 - x \rvert^a \quad a \geq 1$	χ-divergence of order a (Vajda) and Total Variation if $a = 1$ (Saks)

Divergences can be used in both hypotheses testing and estimation (see, e.g. Pardo, 2006), here we consider hypotheses testing problems.

For a given sample of n i.i.d. observations $X_1, \ldots, X_n$ with common density $f(x, \theta)$, under standard regularity assumptions on the model and on ϕ, the standard result is that, under $H_0 : \theta = \theta_0$

$$2n\mathbf{D}_\phi(\hat{\theta}_n, \theta_0) \Rightarrow \chi^2_d$$

with

$$\mathbf{D}_\phi(\hat{\theta}_n, \theta_0) = \frac{1}{n}\sum_{i=1}^{n} \phi\left(\frac{f(X_i, \hat{\theta}_n)}{f(X_i, \theta_0)}\right)$$

where $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ is $\sqrt{n}$-consistent and asymptotically gaussian estimator of θ and d is the dimension of θ

We consider the parametric family of ergodic diffusion process

$$\mathrm{d}X_t = b(\theta_1, X_t)\mathrm{d}t + \sigma(\theta_2, X_t)\mathrm{d}W_t, \quad X_0 = x_0$$

$\theta = (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2 = \Theta$, where $\Theta_1 \subset R^p$ and $\Theta_2 \subset R^q$

b is an $\mathbb{R}^d$-valued function defined on $\mathbb{R}^d \times \Theta_1$, σ is an $\mathbb{R}^d \times \mathbb{R}^d$ valued function defined on $\mathbb{R}^d \times \Theta_2$.

W_t is a d-dimensional Wiener process and x_0 is a deterministic initial value

The process X_t is observed at discrete times $t_i = i\Delta_n, i = 0, 1, 2, ..., n$, where Δ_n is the length of the steps. We denote the observations by $\mathbf{X}_n := \{X_i = X_{t_i}\}_{0\leq i\leq n}$.

The asymptotic is $\Delta_n \to 0, n\Delta_n \to \infty$ and $n\Delta_n^2 \to 0$ as $n \to \infty$.

The pseudo ϕ-divergence

We want to consider hypotheses testing via ϕ-divergences based on discrete time observations from X. We cannot mimic the i.i.d. case, and something like this

$$\frac{1}{n}\sum_{i=1}^{n}\phi\left(\frac{p(X_i,\hat{\theta}_n|X_{i-1})}{p(X_i,\theta_0|X_{i-1})}\right)$$

will just estimate the divergence between invariant measures (with $p(x,\theta|y)$ the transition density).

We want to consider hypotheses testing via ϕ-divergences based on discrete time observations from X. We cannot mimic the i.i.d. case, and something like this

$$\frac{1}{n}\sum_{i=1}^{n}\phi\left(\frac{p(X_i,\hat{\theta}_n|X_{i-1})}{p(X_i,\theta_0|X_{i-1})}\right)$$

will just estimate the divergence between invariant measures (with $p(x,\theta|y)$ the transition density). So we introduce pseudo ϕ-divergences simply defined as follows

$$\mathbb{D}_\phi(\tilde{\theta}_n,\theta_0)=\phi\left(\frac{f_n(\mathbf{X}_n,\tilde{\theta}_n)}{f_n(\mathbf{X}_n,\theta_0)}\right)$$

to test $H_0:\theta=\theta_0$ against $H_1:\theta\neq\theta_0$ where $\tilde{\theta}_n$ is some consistent and asymptotically gaussian estimator of θ and $f_n(\cdot,\theta)$ is an approximated likelihood for $\mathbf{X}_n$

Before stating the result, which differs from the i.i.d. case, we remind few results for divergences and diffusion processes.

Before stating the result, which differs from the i.i.d. case, we remind few results for divergences and diffusion processes.

For continuous time observations from diffusion processes, Vajda (1990) considered the model

$$\mathrm{d}X(t) = -b(t)X_t\mathrm{d}t + \sigma(t)\mathrm{d}W_t$$

and derived explicit formulas for the Rényi divergence

Before stating the result, which differs from the i.i.d. case, we remind few results for divergences and diffusion processes.

For continuous time observations from diffusion processes, Vajda (1990) considered the model

$$\mathrm{d}X(t) = -b(t)X_t\mathrm{d}t + \sigma(t)\mathrm{d}W_t$$

and derived explicit formulas for the Rényi divergence

Küchler and Sørensen (1997) and Morales *et al.* (2004) contain several results on the generalized likelihood ratio test statistics and Rényi statistics for exponential families of diffusions

Before stating the result, which differs from the i.i.d. case, we remind few results for divergences and diffusion processes.

For continuous time observations from diffusion processes, Vajda (1990) considered the model

$$\mathrm{d}X(t) = -b(t)X_t\mathrm{d}t + \sigma(t)\mathrm{d}W_t$$

and derived explicit formulas for the Rényi divergence

Küchler and Sørensen (1997) and Morales *et al.* (2004) contain several results on the generalized likelihood ratio test statistics and Rényi statistics for exponential families of diffusions

Explicit derivations of the Rényi information on the invariant law of ergodic diffusion processes have been presented in De Gregorio and I. (2008)

For continuous time small diffusion processes f-unbiased information criteria have been derived in Uchida and Yoshida (2004) by means of Malliavin calculus. For mixing processes see Uchida and Yoshida (2001)

For continuous time small diffusion processes f-unbiased information criteria have been derived in Uchida and Yoshida (2004) by means of Malliavin calculus. For mixing processes see Uchida and Yoshida (2001)

Rivas *et al.* (2005) derived Rényi divergences for discrete time observations from the model $\mathrm{d}X_t = a\mathrm{d}t + b\mathrm{d}W_t$ where a and b are two scalars

For continuous time small diffusion processes f-unbiased information criteria have been derived in Uchida and Yoshida (2004) by means of Malliavin calculus. For mixing processes see Uchida and Yoshida (2001)

Rivas *et al.* (2005) derived Rényi divergences for discrete time observations from the model $\mathrm{d}X_t = a\mathrm{d}t + b\mathrm{d}W_t$ where a and b are two scalars

Akaike Information Criteria for discretely observed diffusion processes was derived by Uchida and Yoshida (2005)

Kutoyants (2004) and Dachian and Kutoyants (2008) consider the problem of testing statistical hypotheses for ergodic diffusion models in continuous time

Review: hypotheses testing

Kutoyants (2004) and Dachian and Kutoyants (2008) consider the problem of testing statistical hypotheses for ergodic diffusion models in continuous time

Kutoyants (1984) and I. and Kutoyants (2001) consider parametric and semiparametric hypotheses testing for small diffusion processes

Review: hypotheses testing

Kutoyants (2004) and Dachian and Kutoyants (2008) consider the problem of testing statistical hypotheses for ergodic diffusion models in continuous time

Kutoyants (1984) and I. and Kutoyants (2001) consider parametric and semiparametric hypotheses testing for small diffusion processes

Negri and Nishiyama (2008) propose a non parametric test based on score marked empirical process for continuous time observations of ergodic diffusions and Masuda *et al.* (2008) analyzed the discrete time case. Lee and Wee (2008) considered the parametric version of this test for a simplified ergodic model. Negri and Nishiyama (2007) studied the same test for continuous and discrete time observations from small diffusion processes.

Nonparametric signal detection with small error type I and II error probabilities (Ermakov, 2003, today!)

Aït-Sahalia (1996), Giet and Lubrano (2008) and Chen *et al.* (2008) proposed tests based on the several distances between parametric and nonparametric estimation of the invariant density of discretely observed ergodic diffusion processes

Review: hypotheses testing

Aït-Sahalia (1996), Giet and Lubrano (2008) and Chen *et al.* (2008) proposed tests based on the several distances between parametric and nonparametric estimation of the invariant density of discretely observed ergodic diffusion processes

(Up to our knowledge) No other option exists for parametric hypotheses testing based on divergences for discretely observed diffusions processes, and this was the motivation for this work

Consider again the ϕ-divergence

$$D_\phi(\theta, \theta_0) = E_{\theta_0} \phi \left(\frac{p(X, \theta)}{p(X, \theta_0)} \right)$$

where $p(X, \theta)$ is the likelihood of the process X under θ.

Let $\phi(\cdot)$ be such that $\phi(1) = 0$ and denote $C_\phi = \phi'(1)$ and $K_\phi = \phi''(1)$.

Consider again the ϕ-divergence

$$D_\phi(\theta, \theta_0) = E_{\theta_0} \phi \left(\frac{p(X, \theta)}{p(X, \theta_0)} \right)$$

where $p(X, \theta)$ is the likelihood of the process X under θ.

Let $\phi(\cdot)$ be such that $\phi(1) = 0$ and denote $C_\phi = \phi'(1)$ and $K_\phi = \phi''(1)$.

In order to get additional properties, in the i.i.d. case $\phi(x)$ is assumed to be convex or decreasing in $x \in (0, 1)$ and increasing for $x > 1$. These conditions are very convenient in the presence of exponential families. We do not ask for these conditions in our framework.

Back to our test statistic

Given the observations $\mathbf{X}_n := \{X_i = X_{t_i}\}_{0 \le i \le n}$ want to test $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ using the pseudo ϕ-divergence

$$\mathbb{D}_\phi(\tilde{\theta}_n, \theta_0) = \phi\left(\frac{f_n(\mathbf{X}_n, \hat{\theta}_n)}{f_n(\mathbf{X}_n, \theta_0)}\right)$$

with $f_n(\cdot, \theta)$ is the local gaussian approximation of the likelihood of $\mathbf{X}_n$.

$$\begin{aligned} f_n(\mathbf{X}_n, \theta) \quad = \quad & \prod_{i=1}^n \frac{1}{(2\pi\Delta_n)^{d/2} \text{det}(\Sigma(\theta_2, X_{i-1}))^{1/2}} \\ & \times \exp\left\{ -\frac{1}{2\Delta_n} \sum_{i=1}^n [X_i - X_{i-1} - \Delta_n b(\theta_1, X_{i-1})]' \Sigma(\theta_2, X_{i-1})^{-1} \right. \\ & \left. \times [X_i - X_{i-1} - \Delta_n b(\theta_1, X_{i-1})] \right\} \end{aligned}$$

where $\Sigma(\theta, x) = \sigma(\theta, x)\sigma'(\theta, x)$

Back to our test statistic

From Yoshida (1992, and 2006 PLD paper) by quasi-likelihood analysis or Kessler (1997), we get an estimator $\hat{\theta}_n$, such that

$$\Gamma^{-1/2}(\tilde{\theta}_n - \theta_0) \xrightarrow{d} N(0, \mathcal{I}(\theta_0)^{-1})$$

with $\mathcal{I}(\theta_0)$ the Fisher information matrix (positive definite and invertible at θ_0) for the model and Γ the $(p+q) \times (p+q)$ matrix

$$\Gamma = \begin{pmatrix} \frac{1}{n\Delta_n} I_p & 0 \\ 0 & \frac{1}{n} I_q \end{pmatrix}$$

with I_p is the $p \times p$ identity matrix.

Any other estimator with the above properties will also work (e.g. Bayesian type estimator).

We discuss other regularity conditions just after next theorem because they apply to a very general setting, not just diffusion processes.

Main result [$C_\phi = \phi'(1)$, $K_\phi = \phi''(1)$]

Theorem 1: Under $H_0 : \theta = \theta_0$ and the asymptotic $n\Delta_n^2 \to 0$, $\Delta_n \to 0$, $n\Delta_n = T \to \infty$ the pseudo ϕ-divergence test statistics is such that

$$\mathbb{D}_\phi(\hat{\theta}_n, \theta_0) \xrightarrow{d} \frac{1}{2}\zeta$$

where $\zeta = (C_\phi \xi_{p+q} + (C_\phi + K_\phi)\xi_{p+q}^2)$ and $\xi_{p+q} \sim \chi_{p+q}^2$

Theorem 1: Under $H_0 : \theta = \theta_0$ and the asymptotic $n\Delta_n^2 \to 0$, $\Delta_n \to 0$, $n\Delta_n = T \to \infty$ the pseudo ϕ-divergence test statistics is such that

$$\mathbb{D}_\phi(\hat{\theta}_n, \theta_0) \xrightarrow{d} \frac{1}{2}\zeta$$

where $\zeta = (C_\phi \xi_{p+q} + (C_\phi + K_\phi)\xi_{p+q}^2)$ and $\xi_{p+q} \sim \chi_{p+q}^2$

To test $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ we use the test statistics

$$T_\phi = \left| \mathbb{D}_\phi(\hat{\theta}_n, \theta_0) \right|$$

and by c.m.t. we also have the limiting distribution.

The test rejects H_0 if $T_\phi > c_\alpha$, where α is the level of the test and

$$1 - \alpha = P_{\theta_0}(T_\phi < c_\alpha) = P\left(\frac{1}{2}|\zeta| < c_\alpha\right)$$

Important remark

Remind that, in the i.i.d. case, with true ϕ-divergences, we have

$$2n\mathbf{D}_\phi(\hat{\theta}_n, \theta_0) \Rightarrow \chi^2_d$$

i.e. the limiting distribution does not depend on the ϕ used to define the divergence

With pseudo ϕ-divergences we get

$$\mathbb{D}_\phi(\hat{\theta}_n, \theta_0) \xrightarrow{d} \frac{1}{2}\zeta$$

where $\zeta = (C_\phi \xi_{p+q} + (C_\phi + K_\phi)\xi^2_{p+q})$ and $\xi_{p+q} \sim \chi^2_{p+q}$, i.e. the limiting distribution depends on the function ϕ.

Can we try to characterize this limit?

The power function of the test

We can study the power function of the test analytically under contiguous alternatives of the type $H_1 : \theta = \theta_0 + \varphi(n)h$, where $h : \theta = \theta_0 + \varphi(n)h \in \Theta$, with $\varphi(n)$ the appropriate rate.

The power function of the test

We can study the power function of the test analytically under contiguous alternatives of the type $H_1 : \theta = \theta_0 + \varphi(n)h$, where $h : \theta = \theta_0 + \varphi(n)h \in \Theta$, with $\varphi(n)$ the appropriate rate.

Theorem 2:

$$\beta_\phi(\theta) \cong 1 - \mathbf{F}_\delta\left(c_\alpha\right),$$

where $\mathbf{F}_\delta(\cdot)$ is the cumulative function $\frac{1}{2}|\zeta_\delta|$ with of the random variable

$$\zeta_\delta = C_\phi \xi_{p+q}(\delta) + (C_\phi + K_\phi)\xi^2_{p+q}(\delta)$$

with $\xi_{p+q}(\delta)$ is a noncentral chi-square random variable with $p+q$ degrees of freedom and noncentrality parameter $\delta = h'\mathcal{I}(\theta_0)h$.

Example: behaviour of limiting poser function

We compare several ϕ functions. For example, the α-divergences, i.e. such that

$$\phi_\alpha(x) = 4\frac{\left(1 - x^{\frac{1+\alpha}{2}}\right)}{1-\alpha^2}, \quad -1 < \alpha < 1, \quad C_{\phi_\alpha} = \frac{2}{\alpha - 1}, \quad K_{\phi_\alpha} = 1$$

so the power function depends on α as well. For power divergences with

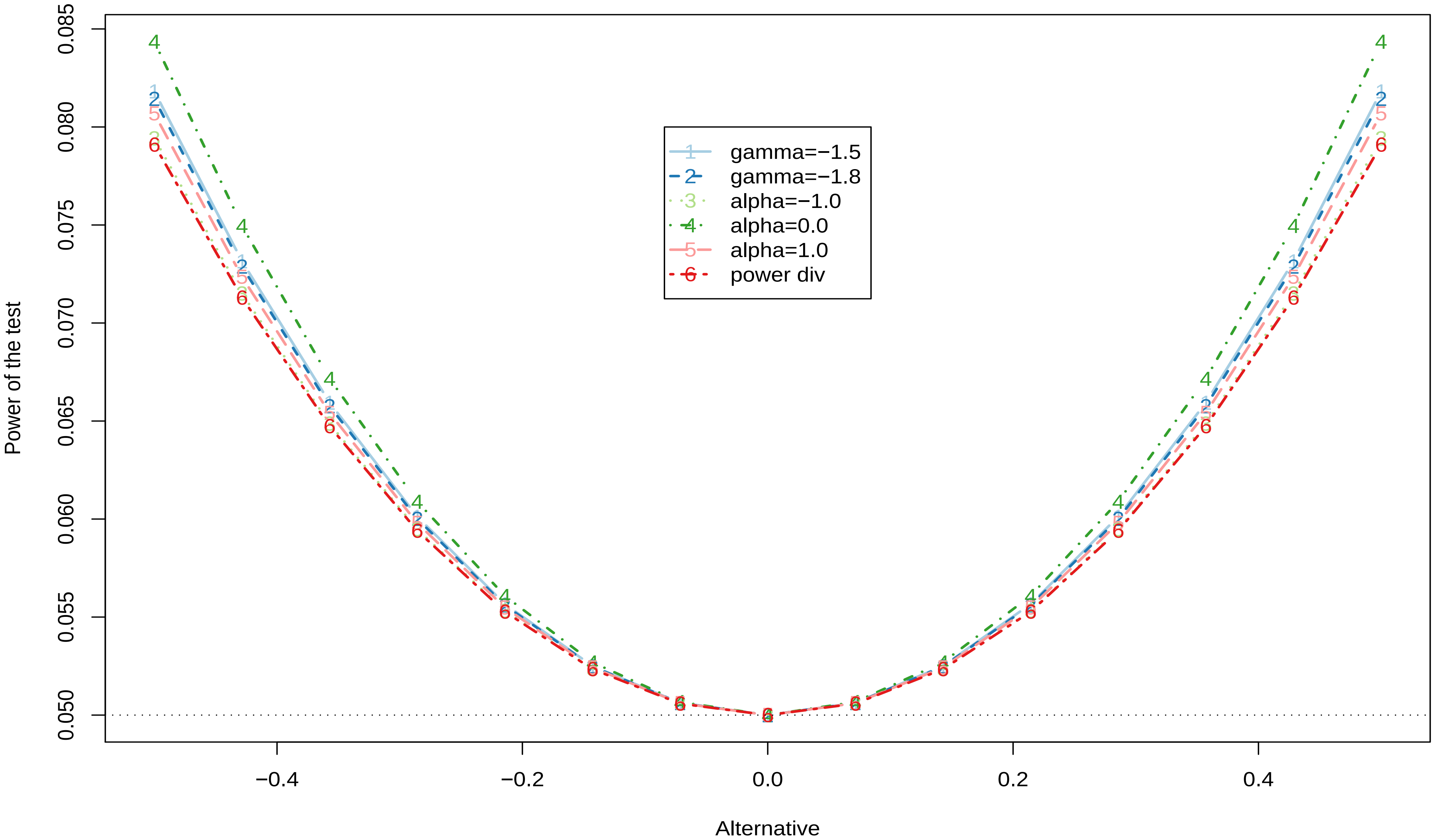
$$\phi_\lambda(x) = \frac{x^{1+\lambda} - \lambda(x-1) - x}{\lambda(\lambda+1)}, \quad \lambda \neq 0, 1 \quad C_{\phi_\lambda} = 0, \quad K_{\phi_\lambda} = 1$$

the limit does not depend on the parameter λ, and the following slight modification

$$\phi_\gamma(x) = \frac{x^{2+\gamma} - \gamma(x-1) - x}{\gamma(\gamma+1)}, \quad \gamma \neq -1, 0 \quad C_{\phi_\gamma} = \frac{1}{\gamma(\gamma+1)}, \quad K_{\phi_\gamma} = \frac{2+\gamma}{\gamma}$$

for which the limit distribution depends on both C_ϕ and K_ϕ.

Asymptotic power under contiguous alternatives



Emprical power under contiguous alternatives

For finite sample size n the power function is more difficult to analyze and the ordering of the ϕ functions is not so clear.

We consider a simple simulation study to show what can happen. We consider a 1-dimensional diffusion process

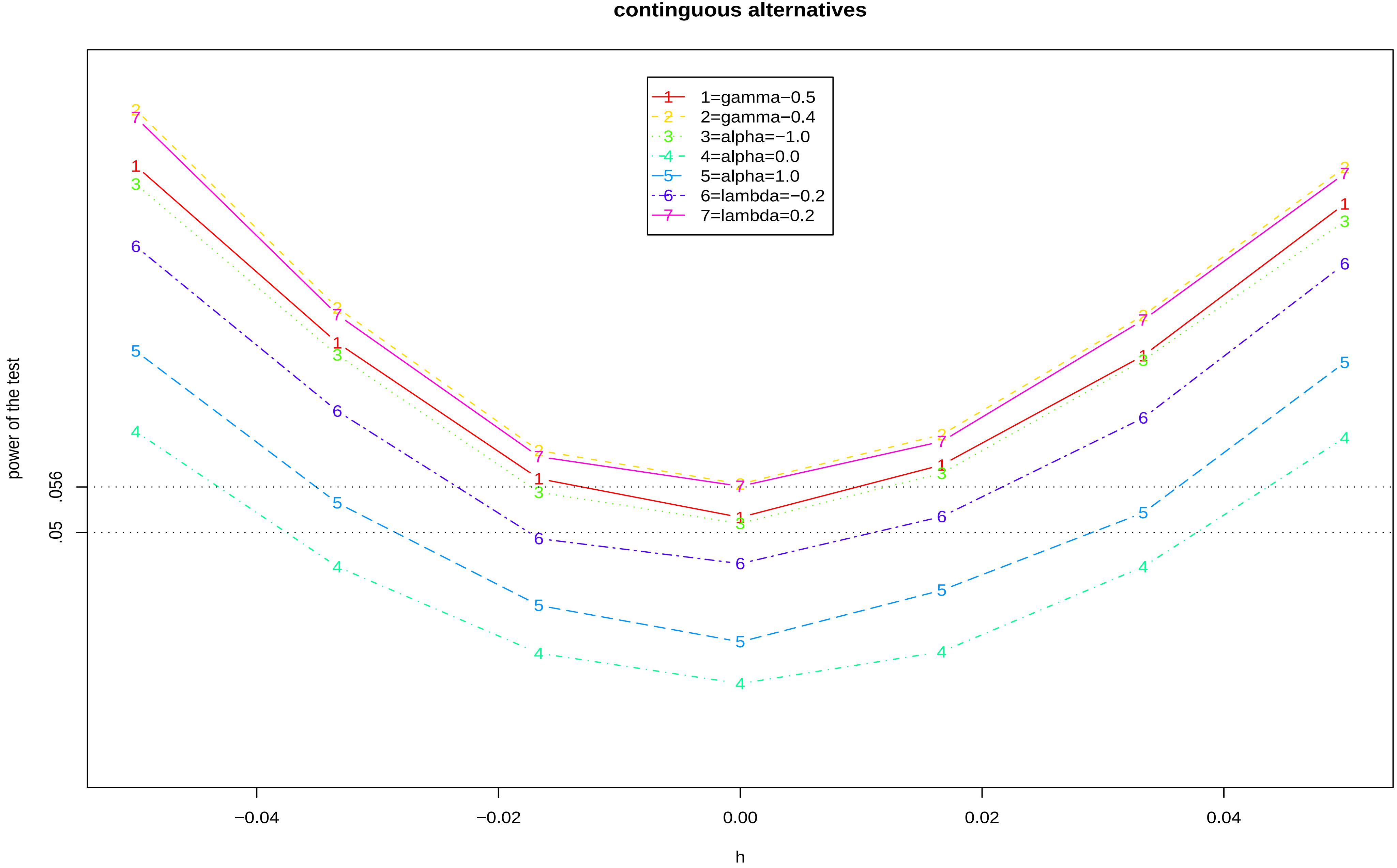
$$\mathrm{d}X_t = (1 - 0.2X_t)\mathrm{d}t + \sqrt{\theta}\mathrm{d}W_t$$

for $\theta_0 = 0.8$ against $\theta_1 = \theta_0 + \frac{h}{\sqrt{n}}$ for extremely small sample size $n = 50$ for $h \in [-0.5, 0.5]$, i.e. $\theta_1 \in [0.73, 0.87]$.

$M = 10,000$ Monte Carlo replications to obtain next picture

From that picture, it emerges that we need to investigate asymptotic expansion of the power function to draw clear conclusions for small n

Empirical power. Sample size $n = 50$



Idea of the proof

The proof is obtained by means of the δ-method up to second order.

These convergence results are needed to prove the theorems

$$\Gamma^{\frac{1}{2}} \nabla_\theta \log f_n(\mathbf{X}_n, \theta_0) \xrightarrow{p} N(0, \mathcal{I}(\theta_0))$$

$$\Gamma^{\frac{1}{2}} \nabla_\theta^2 \log f_n(\mathbf{X}_n, \theta_0) \Gamma^{\frac{1}{2}} \xrightarrow{p} -\mathcal{I}(\theta_0)$$

and to get the power function in explicit form

$$\Gamma \nabla_\theta^3 \log f_n(\mathbf{X}_n, \theta_0) \rightarrow const$$

Idea of the proof

The proof is obtained by means of the δ-method up to second order.

These convergence results are needed to prove the theorems

$$\Gamma^{\frac{1}{2}} \nabla_\theta \log f_n(\mathbf{X}_n, \theta_0) \xrightarrow{p} N(0, \mathcal{I}(\theta_0))$$

$$\Gamma^{\frac{1}{2}} \nabla^2_\theta \log f_n(\mathbf{X}_n, \theta_0) \Gamma^{\frac{1}{2}} \xrightarrow{p} -\mathcal{I}(\theta_0)$$

and to get the power function in explicit form

$$\Gamma \nabla^3_\theta \log f_n(\mathbf{X}_n, \theta_0) \to const$$

Actually, the result hold for any (approximation of the) likelihood f_n and appropriate estimator provided that the above convergences can be proved. So, it works for i.i.d. case as well.

The package `sde` for the `R` statistical environment is freely available at `http://cran.R-Project.org`.

It contains the function `sdeDiv` which implements the ϕ-divergence test statistics.

Software: example of use

We consider the model

$$dX_t = (\theta_{i1} - \theta_{i2} X_t)dt + \theta_{i3}\sqrt{X_t}dW_t, \quad i = 0, 1$$

with (as, in Pritsker, 1998, and Chen *et al.*, 2008)

$$\theta_0 = (0.0807, 0.8922, 0.1809)$$
$$\theta_1 = (0.0403, 0.8922, 0.1279)$$

```
theta0 <- c(0.0807, 0.8922, 0.1809)
theta1 <- c(0.0403, 0.8922, 0.1279)
```

We simulate under $H_1 : \theta = \theta_1$ and test for $H_0 : \theta = \theta_0$

```
set.seed(123)
X <- sde.sim(X0=rsCIR(1, theta1), N=5000, delta=1e-3,  model="CIR", theta=theta1)
```

after setting up model description

```
b <- function(x,theta) theta[1]-theta[2]*x  # drift coefficient
b.x <- function(x,theta)  -theta[2]
s <- function(x,theta) theta[3]*sqrt(x) # diffusion coefficient
s.x <- function(x,theta) theta[3]/(2*sqrt(x))
s.xx <- function(x,theta) -theta[3]/(4*x^1.5)
```

we choose the power divergences

```
lambda <- -1.75
myphi <-  expression((x^(lambda+1) -x -lambda*(x-1))/(lambda*(lambda+1)))
```

$$\phi(x) = \frac{x^{\lambda+1} - x - \lambda(x-1)}{\lambda(\lambda+1)}$$

We run the test. Should reject H_0

```
sdeDiv(X=X, theta0 = theta0, phi = myphi, b=b, s=s, b.x=b.x, s.x=s.x, s.xx=s.xx,
       method="L-BFGS-B", lower=rep(1e-3,3), guess=c(1,1,1))

estimated parameters
0.04041047 1.298524 0.1290066

Testing H0 against H1

H0: 0.0807 0.8922 0.1809
H1: 0.04041047 1.298524 0.1290066

Divergence statistic: 2.8492e+151 (p-value=0)
Likelihood ratio test statistic: 930.69 (p-value=1.9486e-201)
```

Software: example of use

We run the test. Should reject H_0

```
sdeDiv(X=X, theta0 = theta0, phi = myphi, b=b, s=s, b.x=b.x, s.x=s.x, s.xx=s.xx,
       method="L-BFGS-B", lower=rep(1e-3,3), guess=c(1,1,1))

estimated parameters
0.04041047 1.298524 0.1290066

Testing H0 against H1

H0: 0.0807 0.8922 0.1809
H1: 0.04041047 1.298524 0.1290066

Divergence statistic: 2.8492e+151 (p-value=0)
Likelihood ratio test statistic: 930.69 (p-value=1.9486e-201)
```

H_0 successfully rejected! Both by power-divergence and GLRT.

Software: example of use

Now we run the test for $H_0 = H_1$, should not reject

```
sdeDiv(X=X, theta0 = theta1, phi = myphi, b=b, s=s, b.x=b.x, s.x=s.x, s.xx=s.xx,
       method="L-BFGS-B", lower=rep(1e-3,3), guess=c(1,1,1))

estimated parameters
0.04041047 1.298524 0.1290066

Testing H0 against H1

H0: 0.0403 0.8922 0.1279
H1: 0.04041047 1.298524 0.1290066

Divergence statistic: 8.7511 (p-value=0.24091)
Likelihood ratio test statistic: 6.883 (p-value=0.075723)
```

Software: example of use

Now we run the test for $H_0 = H_1$, should not reject

```
sdeDiv(X=X, theta0 = theta1, phi = myphi, b=b, s=s, b.x=b.x, s.x=s.x, s.xx=s.xx,
       method="L-BFGS-B", lower=rep(1e-3,3), guess=c(1,1,1))

estimated parameters
0.04041047 1.298524 0.1290066

Testing H0 against H1

H0: 0.0403 0.8922 0.1279
H1: 0.04041047 1.298524 0.1290066

Divergence statistic: 8.7511 (p-value=0.24091)
Likelihood ratio test statistic: 6.883 (p-value=0.075723)
```

clearly H_0 not rejected at 5% by power divergences. Not rejected also by GLRT with suspect p-value

In the i.i.d. case Jager and Wellner (2007) considered the statistics

$$S_n(\phi) = \sup_x K_\phi(\hat{F}_n(x), F(x)) \quad \text{and} \quad T_n(\phi) = \int_0^1 K_\phi(\hat{F}_n(x), F(x))\mathrm{d}x$$

where $\hat{F}_n$ is the e.d.f. and

$$K_\phi(u,v) = v\phi(u/v) + (1-v)\phi((1-u)/(1-v))$$

similar problem can be considered for the e.d.f. $\hat{F}_T$ of ergodic diffusions: still working on this.

Relevant references

Aït-Sahalia, Y. (1996) Testing continuous-time models of the spot interest rate, *Rev. Financial Stud.*,

Amari, S. (1985) *Differential-Geometrical Methods in Statistics*, Lecture Notes in Statist., Vol. **28**, New York, Springer.

Bhattacharyya, A. (1946) On some analogues to the amount of information and their uses in statistical estimation, *Sankhya*, **8**, 1-14.

Beran, R.J. (1977) Minimum Hellinger estimates for parametric models, *Annals of Statistics*, **5**, 445-463.

Chen, S.X., Gao, J., Cheng, Y.T. (2008) A test for model specification of diffusion processes, *Ann. Stat.*, **36**(1), 167-198.

Cressie, N., Read, T.R.C. (1984) Multinomial goodness of fit tests, *J. Roy. Statist. Soc. Ser B*, **46**, 440-464.

Csiszár, I. (1963) Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. Publ. Math. Inst. Hungar. Acad. Sci. Ser. A **8**, 85-108.

Csiszár, I. (1967) On topological properties of f-divergences, *Studia Scientific Mathematicae Hungarian*, **2**, 329-339.

Dachian, S., Kutoyants, Yu. A. (2008) On the goodness-of-t tests for some continuous time processes, in *Statistical Models and Methods for Biomedical and Technical Systems*, 395-413. Vonta, F., Nikulin, M., Limnios, N. and Huber-Carol C. (Eds), Birkhuser, Boston.

De Gregorio, A., Iacus, S.M. (2008) Rényi information for ergodic diffusion processes. *Information Sciences*, **179**, 279-291.

Giet, L., Lubrano, M. (2008) A minimum Hellinger distance estimator for stochastic differential equations: An application to statistical inference for continuous time interest rate models, *Computational Statistics and Data Analysis*, **52**, 2945-2965.

Küchler, U., Sørensen, M. (1997) *Exponential Families of Stochastic Processes*, Springer, New York.

Kutoyants, Y. (2004) *Statistical Inference for Ergodic Diffusion Processes*, Springer-Verlag, London.

Iacus, S.M., Kutoyants, Y. (2001) Semiparametric hypotheses testing for dynamical systems with small noise, *Mathematical Methods of Statistics*, **10**(1), 105-120.

Jager, L., Wellner, J.A. (2007) Goodness-of-fit tests via phi-divergences, *Annals of Statistics*, **35**(5), 2018-2053.

Relevant references

Lee, S., Wee, I.-S. (2008) Residual emprical process for diffusion processes, *Journal of Korean Mathematical Society*, **45**(3), 2008.

Liese, F., Vajda, I. (1987) *Convex Statistical Distances*, Tuebner, Leipzig.

Masuda, H., Negri, I., Nishiyama, Y. (2008) Goodness of fit test for ergodic diffusions by discrete time observations: an innovation martingale approach, *Research Memorandum 1069, Inst. Statist. Math.*, Tokyo.

Morales, D., Pardo, L., Pardo,, M.C., Vajda, I. (2004) Rényi statistics in directed families of exponential experiments, *Statistics*, **38**(2), 133-147.

Negri, I., Nishiyama, Y. (2008) Goodness of fit test for ergodic diffusion processes, to appear in *Ann. Inst. Statist. Math.*.

Negri, I. and Nishiyama, Y. (2007). Goodness of t test for small diffusions based on discrete observations, *Research Memorandum 1054, Inst. Statist. Math.*, Tokyo.

Pardo, L. (2006) *Statistical Inference Based on Divergence Measures*, Chapman & Hall/CRC, London.

Pritsker, M. (1998) Nonparametric density estimation and tests of continuous time interest rate models, *Rev. financial Studies*, **11**, 449-487.

Rényi, A. (1961) On measures of entropy and information, *Proceedings of the Fourth Berkeley Symposium on Probability and Mathematical Statistics*, Vol 1, University of California, Berkeley, 547-461.

Rivas, M.J., Santos, M.T., Morales, D. (2005) Rényi test statistics for partially observed diffusion processes, *Journal of Statistical Planning and Inference*, **127**, 91-102.

Simpson, D.G. (1989) Hellinger deviance tests: Efficinecy, breakdown points, and examples, *J. Amer. Statist. Assoc.*, **84**, 107-113.

Uchida, M., Yoshida, N. (2001) Information Criteria in Model Selection for Mixing Processes, *Statistical Inference for Stochastic Processes*, **4**, 73-98.

Uchida, M., Yoshida, N. (2004) Information Criteria for Small Diffusions via the Theory of Malliavin-Watanabe, *Statistical Inference for Stochastic Processes*, **7**, 35-67.

Uchida, M., Yoshida, N. (2005) AIC for ergodic diffusion processes from discrete observations, preprint MHF 2005-12, march 2005, *Faculty of Mathematics, Kyushu University*, Fukuoka, Japan.

Yoshida, N. (1992) Estimation for diffusion processes from discrete observation, *J. Multivar. Anal.*, **41**(2), 220-242.