Divergence Test Statistics for Discretely Observed Diffusion Processes

S.M. Iacus
University of Milan

(joint work with A. De Gregorio, UniRoma1)

X Workshop on Quantitative Finance, Polimi 29-Jan-09
Divergences are measure of discrepancy between statistical models.

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

The measures \( P_\theta \) are assumed to have densities \( p_\theta \) w.r.t some common dominating measure \( \mu \) on \( \mathcal{X} \)

\[
p(x, \theta) = \frac{dP_\theta}{d\mu}(x)
\]
$\phi$-Divergences

$\phi$-divergences are defined as follows

$$D_{\phi}(\theta, \theta_0) = \int_X p(\theta_0, x) \phi \left( \frac{p(\theta, x)}{p(\theta_0, x)} \right) \mu(dx)$$

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

and the minimal requirement on the function $\phi$ is: $\phi(1) = 0$
The $\phi$-divergences were 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 $\alpha$-divergences (Csiszar, 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 $\alpha$-divergences include some special cases.

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

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

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 $\alpha$-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

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

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

[known as Bhattacharyya (1946) divergence], otherwise

$$D_{\alpha}(\theta_0, \theta) = D_{1-\alpha} (\theta, \theta_0)$$
Examples: power divergences

The transformation

\[ \psi(R_\alpha) = \frac{\exp\{(\alpha - 1)R_\alpha - 1\}}{1 - \alpha} \]

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

Power divergences \( D_{\phi_\lambda} \) can be obtained directly from the \( \phi \)-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 $\phi$-divergences (see Pardo, 2006)

<table>
<thead>
<tr>
<th>$\phi(x)$ with $x = p(\theta, \cdot)/p(\theta_0, \cdot)$</th>
<th>Divergence</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x \log x - x + 1$</td>
<td>Kullback-Leibler</td>
</tr>
<tr>
<td>$- \log x + -1$</td>
<td>Minimum Discrimination Information</td>
</tr>
<tr>
<td>$(x - 1) \log x$</td>
<td>$J$-divergence</td>
</tr>
<tr>
<td>$\frac{1}{2} (x - 1)^2$</td>
<td>Pearson, Kagan</td>
</tr>
<tr>
<td>$\frac{(x-1)^2}{(x+1)^2}$</td>
<td>Balakrishnan &amp; Sanghvi</td>
</tr>
<tr>
<td>$-x^s + s(x-1)+1 \frac{1}{1-s}$, $s \neq 1$</td>
<td>Rathie &amp; Kannappan</td>
</tr>
<tr>
<td>$\frac{1-x}{2} - \left( \frac{1+x-r}{2} \right)^{-1/r}$, $r &gt; 0$</td>
<td>Harmonic mean (Mathai &amp; Rathie)</td>
</tr>
<tr>
<td>$\frac{(1-x)^2}{2(a+(1-a)x)}$, $0 \leq a \leq 1$</td>
<td>Rukhin</td>
</tr>
<tr>
<td>$ax \log x - (ax+1-a) \log(ax+1-a) \frac{a(1-a)}{a} = 0, 1$</td>
<td>Lin</td>
</tr>
<tr>
<td>$\frac{x^{\lambda+1} - x - \lambda(x-1)}{\lambda(\lambda+1)}$, $\lambda \neq 0, -1$</td>
<td>Cressie &amp; Read</td>
</tr>
<tr>
<td>$</td>
<td>1 - x^a</td>
</tr>
<tr>
<td>$</td>
<td>1 - x</td>
</tr>
</tbody>
</table>
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 \( X_1, \ldots, X_n \), under standard regularity assumptions on the model and on \( \phi \), the standard result is that, under \( H_0 : \theta = \theta_0 \)

\[
2nD_\phi(\hat{\theta}_n, \theta_0) \Rightarrow \chi^2_d
\]

where \( \hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n) \) is \( \sqrt{n} \)-consistent and asymptotically gaussian estimator of \( \theta \) and \( d \) is the dimension of \( \theta \).
We consider the parametric family of diffusion process

\[ dX_t = b(\alpha, X_t)dt + \sigma(\beta, X_t)dW_t, \quad X_0 = x_0, \]

\[ \theta = (\alpha, \beta) \in \Theta_\alpha \times \Theta_\beta = \Theta, \text{ where } \Theta_\alpha \subset \mathbb{R}^p \text{ and } \Theta_\beta \subset \mathbb{R}^q. \]

The drift and diffusion coefficients are known up to \( \alpha \) and \( \beta \) and such that the solution of the sde exists and the process is also ergodic.

We will consider hypotheses testing via \( \phi \)-divergences based on discrete time observations from \( X \).
We consider the parametric family of diffusion process

\[ dX_t = b(\alpha, X_t)dt + \sigma(\beta, X_t)dW_t, \quad X_0 = x_0, \]

\[ \theta = (\alpha, \beta) \in \Theta_\alpha \times \Theta_\beta = \Theta, \text{ where } \Theta_\alpha \subset \mathbb{R}^p \text{ and } \Theta_\beta \subset \mathbb{R}^q. \]

The drift and diffusion coefficients are known up to \( \alpha \) and \( \beta \) and such that the solution of the sde exists and the process is also ergodic.

We will consider hypotheses testing via \( \phi \)-divergences based on discrete time observations from \( X \).

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

\[ dX(t) = -b(t)X_t \, dt + \sigma(t) \, dW_t \]

and derived explicit formulas for the Rényi divergence.
For continuous time observations from diffusion processes, Vajda (1990) considered the model

$$dX(t) = -b(t)X_t dt + \sigma(t)dW_t$$

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

\[ dX(t) = -b(t)X_t dt + \sigma(t)dW_t \]

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 in 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 in Uchida and Yoshida (2001)

Rivas et al. (2005) derived Rényi divergences for discrete time observations from the model \( dX_t = adt + bdW_t \) 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 in Uchida and Yoshida (2001)

Rivas et al. (2005) derived Rényi divergences for discrete time observations from the model $dX_t = adt + b dW_t$ where $a$ and $b$ are two scalars

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


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.
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.
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 $\phi$-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 $\theta$.

Let $\phi(\cdot)$ be such that $\phi(1) = 0$. When they exist, define $C_\phi = \phi'(1)$ and $K_\phi = \phi''(1)$. 


Consider again the $\phi$-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 $\theta$.

Let $\phi(\cdot)$ be such that $\phi(1) = 0$. When they exist, define $C_\phi = \phi'(1)$ 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.
The test statistic

We assume that the process $X_t$ is ergodic for every $\theta$ with invariant law $\mu_\theta$. The process $X_t$ is observed at discrete times $t_i = i\Delta_n, i = 0, 1, 2, \ldots, n$, where $\Delta_n$ is the length of the steps. We denote the observations by $X_n := \{X_i = X_{t_i}\}_{0 \leq i \leq n}$.

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

Given $\tilde{\theta}_n$ a consistent and asymptotically gaussian estimator for $\theta$, we propose the following test statistics based on $\phi$-divergence

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

to test $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$.

Here $f_n(\cdot, \theta)$ is an the approximate likelihood of the observed diffusion.
The test statistic

The estimator \( \tilde{\theta}_n \) must be such such that

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

where \( I(\theta_0) \) the Fisher information matrix, positive definite and invertible at \( \theta_0 \)

\[
I(\theta_0) = \begin{pmatrix}
(I_{b}^{k,j}(\theta_0))_{k,j=1,\ldots,p} & 0 \\
0 & (I_{\sigma}^{k,j}(\theta_0))_{k,j=1,\ldots,q}
\end{pmatrix}
\]

with

\[
I_{b}^{k,j}(\theta_0) = \int \frac{1}{\sigma^2(\theta_0, x)} \frac{\partial b(\alpha_0, x)}{\partial \alpha_k} \frac{\partial b(\alpha_0, x)}{\partial \alpha_j} \mu_{\theta_0}(dx)
\]

\[
I_{\sigma}^{k,j}(\theta_0) = 2 \int \frac{1}{\sigma^2(\beta_0, x)} \frac{\partial \sigma(\beta_0, x)}{\partial \beta_k} \frac{\partial \sigma(\beta_0, x)}{\partial \beta_j} \mu_{\theta_0}(dx)
\]

and \( \Gamma \) 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.
Approximation of the density statistics

As in Uchida and Yoshida (2005), consider the following approximation of the likelihood

\[ f_n(\theta) = \exp \{ u_n(\theta) \} , \quad u_n(\theta) = \sum_{k=1}^{n} u(\Delta_n, X_{i-1}, X_i, \theta) \]

where

\[ u(t, x, y, \theta) = -\frac{1}{2} \log(2\pi t) - \log \sigma(y, \beta) - \frac{S^2(x, y, \beta)}{2t} + H(x, y, \theta) + t\tilde{g}(x, y, \theta) , \]

with

\[ S(x, y, \beta) = \int_x^y \frac{\text{d}u}{\sigma(u, \beta)} , \quad H(x, y, \theta) = \int_x^y \frac{B(u, \theta)}{\sigma(u, \beta)} \text{d}u \]

\[ \tilde{g}(x, y, \theta) = -\frac{1}{2} \left\{ C(x, \theta) + C(y, \theta) + \frac{1}{3} B(x, \theta)B(y, \theta) \right\} \]

\[ C(x, \theta) = \frac{1}{2} B^2(x, \theta) + \frac{1}{2} B_x(x, \theta)\sigma(x, \beta) , \quad B(x, \theta) = \frac{b(x, \alpha)}{\sigma(x, \beta)} - \frac{1}{2} \sigma_x(x, \beta) \]
Consistent and asymptotically Gaussian estimator

We consider the approximated maximum likelihood estimator $\hat{\theta}_n$ based on the locally Gaussian approximation (see, e.g. Yoshida, 1992), i.e.

$$\hat{\theta}_n = \arg \max_\theta \ell_n(\theta)$$

with

$$\ell_n(\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi \Delta_n \sigma^2(X_{i-1}, \theta)}} \exp \left\{ -\frac{1}{2} \frac{(X_i - X_{i-1} - b(X_{i-1}, \theta) \Delta_n)^2}{\Delta_n \sigma^2(X_{i-1}, \theta)} \right\}$$

The estimator $\hat{\theta}_n$ satisfies previous convergence assumptions for $\tilde{\theta}_n$. 
Regularity conditions on the process

i) There exists a constant $C$ such that

$$|b(\alpha_0, x) - b(\alpha_0, y)| + |\sigma(\beta_0, x) - \sigma(\beta_0, y)| \leq C|x - y|.$$ 

ii) $\inf_{\beta, x} \sigma^2(\beta, x) > 0$.

iii) The process $X$ is ergodic for every $\theta$ with invariant probability measure $\mu_\theta$. All polynomial moments of $\mu_\theta$ are finite.

iv) For all $m \geq 0$ and for all $\theta$, $\sup_t E|X_t|^m < \infty$.

v) For every $\theta$, the coefficients $b(\alpha, x)$ and $\sigma(\beta, x)$ are five times differentiable with respect to $x$ and the derivatives are polynomial growth in $x$, uniformly in $\theta$.

vi) The coefficients $b(\alpha, x)$ and $\sigma(\beta, x)$ and all their partial derivatives respect to $x$ up to order 2 are three times differentiable respect to $\theta$ for all $x$ in the state space. All derivatives respect to $\theta$ are polynomial growth in $x$, uniformly in $\theta$. 
Let \( i = 0, 1, 2, 3 \) and \( \partial_{\theta}^i \) the partial \( i \)-th derivative with respect to \( \theta \) and similarly for \( x \),

i) \( \partial_{\theta}^i \tilde{h}(x, \theta) = O(|x|^2) \) as \( x \to \infty \).

ii) \( \inf_x \partial_{\theta}^i \tilde{h}(x, \theta) > -\infty \)

iii) \( \sup_{\theta} \sup_x |\partial_{\theta}^i \partial_{x}^5 \tilde{h}(x, \theta)| \leq M < \infty \).

iv) There exists \( \gamma > 0 \) such that for every \( \theta \) and \( j = 1, \ldots, 4 \),
\[
|\partial_{\theta}^i \partial_{x}^j \tilde{B}(x, \theta)| = O(|\tilde{B}(x, \theta)|^\gamma) \text{ as } |x| \to \infty.
\]

When the coefficients \( b(\alpha, x) = b(\alpha_0, x) \) and \( \sigma^2(\beta, x) = \sigma^2(\beta_0, x) \) for \( \mu_{\theta_0} \) a.s. for all \( x \), then \( \alpha = \alpha_0 \) and \( \beta = \beta_0 \).
Main result \[ C_\phi = \phi'(1), \quad K_\phi = \phi''(1) \]

Under \( H_0 : \theta = \theta_0, \ n\Delta_n^2 \to 0, \ \Delta_n \to 0, \ n\Delta_n = T \to \infty \)

i) if \( 0 \neq C_\phi < \infty \) and \( K_\phi = 0 \), then (like in the i.i.d. case)

\[ \mathbb{D}_\phi(\hat{\theta}_n, \theta_0) \xrightarrow{d} C_\phi \chi_{p+q}^2 \]
Main result \[ C_\phi = \phi'(1), \quad K_\phi = \phi''(1) \]

Under \( H_0 : \theta = \theta_0, \ n\Delta_n^2 \to 0, \ \Delta_n \to 0, \ n\Delta_n = T \to \infty \)

i) if \( 0 \neq C_\phi < \infty \) and \( K_\phi = 0 \), then (like in the i.i.d. case)

\[
\mathbb{D}_\phi(\hat{\theta}_n, \theta_0) \xrightarrow{d} C_\phi \chi^2_{p+q}
\]

ii) if \( C_\phi = 0 \) and \( 0 \neq K_\phi < \infty \), then

\[
\mathbb{D}_\phi(\hat{\theta}_n, \theta_0) \xrightarrow{d} \frac{K_\phi}{2} Z_{p+q} \quad \sqrt{Z_{p+q}} \sim \chi^2_{p+q}
\]
Main result \[ [C_\phi = \phi'(1), \; K_\phi = \phi''(1)] \]

Under \( H_0 : \theta = \theta_0, \; n\Delta_n^2 \to 0, \; \Delta_n \to 0, \; n\Delta_n = T \to \infty \)

i) if \( 0 \neq C_\phi < \infty \) and \( K_\phi = 0 \), then (like in the i.i.d. case)
\[
\mathbb{D}_\phi(\hat{\theta}_n, \theta_0) \xrightarrow{d} C_\phi \chi^2_{p+q}
\]

ii) if \( C_\phi = 0 \) and \( 0 \neq K_\phi < \infty \), then
\[
\mathbb{D}_\phi(\hat{\theta}_n, \theta_0) \xrightarrow{d} \frac{K_\phi}{2} Z_{p+q} \quad \sqrt{Z_{p+q}} \sim \chi^2_{p+q}
\]

iii) if \( C_\phi \neq 0 \) and \( K_\phi \neq 0 \) are finite, then
\[
\mathbb{D}_\phi(\hat{\theta}_n, \theta_0) \xrightarrow{d} \frac{1}{2} \left( C_\phi \chi^2_{p+q} + (C_\phi + K_\phi)Z_{p+q} \right)
\]
For example, consider the $\alpha$-divergences

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

and the limit as $\alpha \rightarrow -1$, i.e. the Kullback-Leibler divergence, we have

$$\phi(x) = \lim_{\alpha \rightarrow -1} \phi_\alpha(x) = - \log(x)$$

for which $C_\phi = -1$ and $K_\phi = 1$. Hence

$$\mathbb{D}_{Kull}(\hat{\theta}_n, \theta_0) = \mathbb{D}_\phi(\hat{\theta}_n, \theta_0) \xrightarrow{d} \frac{1}{2} (C_\phi \chi^2_{p+q} + (C_\phi + K_\phi) Z_{p+q})$$

reduces to the standard result of the i.i.d. setting.
Idea of the proof

The proof is obtained by means of the $\delta$-method up to second order.

These lemmas are needed to prove convergence

$$\Gamma^{\frac{1}{2}} \nabla_{\theta} \log \ell_n(X_n, \theta_0) \overset{p}{\to} N(0, \mathcal{I}(\theta_0))$$  \hspace{1cm} \text{(Kessler, 1997)}

$$\Gamma^{\frac{1}{2}} \nabla_{\theta} \log f_n(X_n, \theta_0) = \Gamma^{\frac{1}{2}} \nabla_{\theta} \log \ell_n(X_n, \theta_0) + o_p(1)$$  \hspace{1cm} \text{(Uchida & Yoshida, 2005)}

$$\Gamma^{\frac{1}{2}} \nabla_{\theta}^2 \log f_n(X_n, \theta_0) \Gamma^{\frac{1}{2}} \overset{p}{\to} -\mathcal{I}(\theta_0)$$  \hspace{1cm} \text{(Uchida & Yoshida, 2005)}

So the result hold for any approximation of the likelihood $f_n$ and appropriate estimator provided that the above lemmas can be proved. Similarly for the multidimensional case.
Simulation study

- $\alpha$-divergences

$$D_{\alpha}(\hat{\theta}_n, \theta_0) = \phi_{\alpha} \left( \frac{f_n(X_n, \hat{\theta}_n)}{f_n(X_n, \theta_0)} \right)$$

with $\phi_{\alpha}(x) = 4(1 - x^{\frac{1+\alpha}{2}})/(1 - \alpha^2)$, with $C_{\alpha} = \frac{2}{\alpha - 1}$ and $K_{\phi} = 1$. With $\alpha \in \{-0.99, -0.90, -0.75, -0.50, -0.25, -0.10\}$;

- power-divergences of order $\lambda$

$$D_{\lambda}(\hat{\theta}_n, \theta_0) = \phi_{\lambda} \left( \frac{f_n(X_n, \hat{\theta}_n)}{f_n(X_n, \theta_0)} \right)$$

with $\phi_{\lambda}(x) = (x^{\lambda+1} - x - \lambda(x - 1))/(\lambda(\lambda + 1))$, with $C_{\lambda} = 0$, $K_{\lambda} = 1$. With $\lambda \in \{-0.99, -1.20, -1.50, -1.75, -2.00, -2.50\}$;

- generalized likelihood ratio test statistic

$$D_{\text{log}}(\hat{\theta}_n, \theta_0) = -\log \left( \frac{f_n(X_n, \hat{\theta}_n)}{f_n(X_n, \theta_0)} \right)$$
The parameters of the next models, have been chosen according to Pritsker (1998) and Chen et al. (2008), in particular \( VAS_0 \) corresponds to the model estimated by Aït-Sahalia (1996).

Also null \( H_0 \) and alternative \( H_1 \) hypotheses have been chosen accordingly to Pritsker (1998) and Chen et al. (2008) in order to replicate the framework of their simulation study.
**Model comparisons**

The Vasicek (VAS) model: \[ dX_t = \kappa(\alpha - X_t)dt + \sigma dW_t, \] where, in finance, \( \sigma \) is interpreted as volatility, \( \alpha \) is the long-run equilibrium value of the process and \( \kappa \) is the speed of reversion. Let

\[ \theta_0 = (\kappa_0, \alpha_0, \sigma_0^2) = (0.85837, 0.089102, 0.0021854) \]

we consider three different sets of hypotheses for the parameters

<table>
<thead>
<tr>
<th>model</th>
<th>( \theta = (\kappa, \alpha, \sigma^2) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAS_0</td>
<td>( (\kappa_0, \alpha_0, \sigma_0^2) )</td>
</tr>
<tr>
<td>VAS_1</td>
<td>( (4 \cdot \kappa_0, \alpha_0, 4 \cdot \sigma_0^2) )</td>
</tr>
<tr>
<td>VAS_2</td>
<td>( (\frac{1}{4} \kappa_0, \alpha_0, \frac{1}{4} \cdot \sigma_0^2) )</td>
</tr>
</tbody>
</table>
The Cox-Ingersoll-Ross (CIR) model: \[ dX_t = \kappa(\alpha - X_t)dt + \sigma \sqrt{X_t}dW_t. \]

Let \[ \theta_0 = (\kappa_0, \alpha_0, \sigma_0^2) = (0.89218, 0.09045, 0.032742) \]

we consider different sets of hypotheses for the parameters

<table>
<thead>
<tr>
<th>model</th>
<th>[ \theta = (\kappa, \alpha, \sigma^2) ]</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIR(_0)</td>
<td>[ (\kappa_0, \alpha_0, \sigma_0^2) ]</td>
</tr>
<tr>
<td>CIR(_1)</td>
<td>[ (\frac{1}{2} \cdot \kappa_0, \alpha_0, \frac{1}{2} \cdot \sigma_0^2) ]</td>
</tr>
<tr>
<td>CIR(_2)</td>
<td>[ (\frac{1}{4} \cdot \kappa_0, \alpha_0, \frac{1}{4} \cdot \sigma_0^2) ]</td>
</tr>
</tbody>
</table>

This model has a transition density of \( \chi^2 \)-type, hence local gaussian approximation is less likely to hold for non negligible values of \( \Delta_n \).

The parameters of the above models, have been chosen according to Pritsker (1998) and Chen et al. (2008), in particular VAS\(_0\) corresponds to the model estimated by Aït-Sahalia (1996).
The empirical level of the test is calculated as the number of times the test rejects the null hypothesis under the true model, i.e.

$$\hat{\alpha}_n = \frac{1}{M} \sum_{i=1}^{M} 1\{D_\phi > c_\alpha\} \quad \text{under} \quad H_0$$

where $1_A$ is the indicator function of set $A$, $M = 10,000$ is the number of simulations and $c_\alpha$ is the $(1 - \alpha)%$ quantile of the proper distribution. Similarly we calculate the power of the test under alternative models as

$$\hat{\beta}_n = \frac{1}{M} \sum_{i=1}^{M} 1\{D_\phi > c_\alpha\} \quad \text{under} \quad H_1$$
Vasicek. Alpha-div. of order $a$, $a = -0.99 = \text{GLRT}$

<table>
<thead>
<tr>
<th>model $(\alpha, n)$</th>
<th>$a = -0.99$</th>
<th>$a = -0.90$</th>
<th>$a = -0.75$</th>
<th>$a = -0.50$</th>
<th>$a = -0.25$</th>
<th>$a = -0.10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAS$_0$ (0.01, 50)</td>
<td>0.01</td>
<td>0.10</td>
<td>0.39</td>
<td>0.62</td>
<td>0.73</td>
<td>0.77</td>
</tr>
<tr>
<td>VAS$_1$ (0.01, 50)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VAS$_2$ (0.01, 50)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VAS$_0$ (0.05, 50)</td>
<td>0.04</td>
<td>0.12</td>
<td>0.39</td>
<td>0.62</td>
<td>0.73</td>
<td>0.77</td>
</tr>
<tr>
<td>VAS$_1$ (0.05, 50)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VAS$_2$ (0.05, 50)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VAS$_0$ (0.01, 100)</td>
<td>0.01</td>
<td>0.10</td>
<td>0.39</td>
<td>0.63</td>
<td>0.74</td>
<td>0.78</td>
</tr>
<tr>
<td>VAS$_1$ (0.01, 100)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VAS$_2$ (0.01, 100)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VAS$_0$ (0.05, 100)</td>
<td>0.04</td>
<td>0.11</td>
<td>0.40</td>
<td>0.63</td>
<td>0.74</td>
<td>0.78</td>
</tr>
<tr>
<td>VAS$_1$ (0.05, 100)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VAS$_2$ (0.05, 100)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VAS$_0$ (0.01, 500)</td>
<td>0.02</td>
<td>0.18</td>
<td>0.61</td>
<td>0.83</td>
<td>0.90</td>
<td>0.92</td>
</tr>
<tr>
<td>VAS$_1$ (0.01, 500)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VAS$_2$ (0.01, 500)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VAS$_0$ (0.05, 500)</td>
<td>0.07</td>
<td>0.20</td>
<td>0.61</td>
<td>0.83</td>
<td>0.90</td>
<td>0.92</td>
</tr>
<tr>
<td>VAS$_1$ (0.05, 500)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VAS$_2$ (0.05, 500)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
### Vasicek. Power-div. of order $\lambda$

<table>
<thead>
<tr>
<th>model ($\alpha, n$)</th>
<th>$\lambda = -0.99$</th>
<th>$\lambda = -1.20$</th>
<th>$\lambda = -1.50$</th>
<th>$\lambda = -1.75$</th>
<th>$\lambda = -2.00$</th>
<th>$\lambda = -2.50$</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAS$_0$ (0.01, 50)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>VAS$_1$ (0.01, 50)</td>
<td>0.00</td>
<td>0.99</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VAS$_2$ (0.01, 50)</td>
<td>0.40</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VAS$_0$ (0.05, 50)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.03</td>
<td>0.06</td>
</tr>
<tr>
<td>VAS$_1$ (0.05, 50)</td>
<td>0.67</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VAS$_2$ (0.05, 50)</td>
<td>0.99</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VAS$_0$ (0.01, 100)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>VAS$_1$ (0.01, 100)</td>
<td>0.23</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VAS$_2$ (0.01, 100)</td>
<td>0.88</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VAS$_0$ (0.05, 100)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.03</td>
<td>0.06</td>
</tr>
<tr>
<td>VAS$_1$ (0.05, 100)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VAS$_2$ (0.05, 100)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VAS$_0$ (0.01, 500)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.03</td>
<td>0.08</td>
</tr>
<tr>
<td>VAS$_1$ (0.01, 500)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VAS$_2$ (0.01, 500)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VAS$_0$ (0.05, 500)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.03</td>
<td>0.06</td>
<td>0.12</td>
</tr>
<tr>
<td>VAS$_1$ (0.05, 500)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VAS$_2$ (0.05, 500)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
CIR. Alpha-div. of order $a$, $a = -0.99 = GLRT$

<table>
<thead>
<tr>
<th>model ($\alpha, n$)</th>
<th>$a = -0.99$</th>
<th>$a = -0.90$</th>
<th>$a = -0.75$</th>
<th>$a = -0.50$</th>
<th>$a = -0.25$</th>
<th>$a = -0.10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIR$_0$ (0.01, 50)</td>
<td>0.01</td>
<td>0.14</td>
<td>0.54</td>
<td>0.77</td>
<td>0.85</td>
<td>0.87</td>
</tr>
<tr>
<td>CIR$_1$ (0.01, 50)</td>
<td>0.80</td>
<td>0.98</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CIR$_2$ (0.01, 50)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CIR$_0$ (0.05, 50)</td>
<td>0.05</td>
<td>0.16</td>
<td>0.54</td>
<td>0.77</td>
<td>0.85</td>
<td>0.87</td>
</tr>
<tr>
<td>CIR$_1$ (0.05, 50)</td>
<td>0.94</td>
<td>0.98</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CIR$_2$ (0.05, 50)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CIR$_0$ (0.01, 100)</td>
<td>0.01</td>
<td>0.13</td>
<td>0.49</td>
<td>0.71</td>
<td>0.79</td>
<td>0.82</td>
</tr>
<tr>
<td>CIR$_1$ (0.01, 100)</td>
<td>0.99</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CIR$_2$ (0.01, 100)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CIR$_0$ (0.05, 100)</td>
<td>0.04</td>
<td>0.15</td>
<td>0.49</td>
<td>0.71</td>
<td>0.79</td>
<td>0.82</td>
</tr>
<tr>
<td>CIR$_1$ (0.05, 100)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CIR$_2$ (0.05, 100)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CIR$_0$ (0.01, 500)</td>
<td>0.00</td>
<td>0.06</td>
<td>0.28</td>
<td>0.54</td>
<td>0.69</td>
<td>0.74</td>
</tr>
<tr>
<td>CIR$_1$ (0.01, 500)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CIR$_2$ (0.01, 500)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CIR$_0$ (0.05, 500)</td>
<td>0.02</td>
<td>0.08</td>
<td>0.28</td>
<td>0.54</td>
<td>0.69</td>
<td>0.74</td>
</tr>
<tr>
<td>CIR$_1$ (0.05, 500)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CIR$_2$ (0.05, 500)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
### CIR. Power-div. of order $\lambda$

<table>
<thead>
<tr>
<th>model $(\alpha, n)$</th>
<th>$\lambda = -0.99$</th>
<th>$\lambda = -1.20$</th>
<th>$\lambda = -1.50$</th>
<th>$\lambda = -1.75$</th>
<th>$\lambda = -2.00$</th>
<th>$\lambda = -2.50$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIR$_0$ (0.01, 50)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
<td>0.06</td>
</tr>
<tr>
<td>CIR$_1$ (0.01, 50)</td>
<td>0.00</td>
<td>0.06</td>
<td>0.52</td>
<td>0.75</td>
<td>0.86</td>
<td>0.94</td>
</tr>
<tr>
<td>CIR$_2$ (0.01, 50)</td>
<td>0.00</td>
<td>0.99</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CIR$_0$ (0.05, 50)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>0.04</td>
<td>0.09</td>
</tr>
<tr>
<td>CIR$_1$ (0.05, 50)</td>
<td>0.00</td>
<td>0.23</td>
<td>0.70</td>
<td>0.85</td>
<td>0.92</td>
<td>0.96</td>
</tr>
<tr>
<td>CIR$_2$ (0.05, 50)</td>
<td>0.06</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CIR$_0$ (0.01, 100)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
<td>0.05</td>
</tr>
<tr>
<td>CIR$_1$ (0.01, 100)</td>
<td>0.00</td>
<td>0.56</td>
<td>0.96</td>
<td>0.99</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CIR$_2$ (0.01, 100)</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CIR$_0$ (0.05, 100)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>0.03</td>
<td>0.08</td>
</tr>
<tr>
<td>CIR$_1$ (0.05, 100)</td>
<td>0.00</td>
<td>0.83</td>
<td>0.99</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CIR$_2$ (0.05, 100)</td>
<td>0.97</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CIR$_0$ (0.01, 500)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>CIR$_1$ (0.01, 500)</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CIR$_2$ (0.01, 500)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CIR$_0$ (0.05, 500)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>CIR$_1$ (0.05, 500)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CIR$_2$ (0.05, 500)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
The power divergences are, on average, better than the generalized likelihood ratio test in terms of both empirical level $\hat{\alpha}$ and power $\hat{\beta}$ for the models considered and under the selected alternatives.

The $\alpha$-divergence do not behave very well and only approximate the GLR test at most (i.e. always worse than GLRT).

For the CIR study, all test statistics have, in general, lower power under the alternative CIR$_1$ than under CIR$_2$.

Power divergences are yet the best test statistics in both cases (CIR and VAS), for $\lambda = (-1.20, -1.50, -1.75)$.
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 $\phi$-divergence test statistics.
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)

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

\[
\begin{align*}
\text{theta0} &\leftarrow c(0.0807, 0.8922, 0.1809) \\
\text{theta1} &\leftarrow c(0.0403, 0.8922, 0.1279)
\end{align*}
\]

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

\[
\begin{align*}
\text{set.seed}(123) \\
\text{X} &\leftarrow \text{sde.sim}(X0=\text{rsCIR}(1, \text{theta1}), \text{N}=5000, \text{delta}=1e-3, \text{model}="\text{CIR}" , \text{theta=theta1})
\end{align*}
\]
after setting up model description

\begin{align*}
b &\leftarrow \text{function}(x, \theta) \theta[1] - \theta[2] \times x \quad \# \text{drift coefficient} \\
b.x &\leftarrow \text{function}(x, \theta) \quad -\theta[2] \\
s &\leftarrow \text{function}(x, \theta) \theta[3] \times \sqrt{x} \quad \# \text{diffusion coefficient} \\
s.x &\leftarrow \text{function}(x, \theta) \theta[3] / (2 \times \sqrt{x}) \\
s.xx &\leftarrow \text{function}(x, \theta) \quad -\theta[3] / (4 \times x^{1.5})
\end{align*}

we choose the power divergences

\begin{align*}
\lambda &\leftarrow -1.75 \\
\phi(x) &\leftarrow \text{expression}((x^{(\lambda+1)} - x - \lambda \times (x-1)) / (\lambda \times (\lambda+1)))
\end{align*}

\[
\phi(x) = \frac{x^{\lambda+1} - x - \lambda(x - 1)}{\lambda(\lambda + 1)}
\]
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)
We run the test. Should reject $H_0$

```r
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 $H_0$ against $H_1$

$H_0$: 0.0807 0.8922 0.1809
$H_1$: 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.
Now we run the test for $H_0 = H_1$, should not reject

\[
\text{sdeDiv}(X=X, \theta_0 = \text{thetal}, \phi = \text{myphi}, b=b, s=s, b.x=b.x, s.x=s.x, s.xx=s.xx, \\
\text{method}="L-BFGS-B", \text{lower}=\text{rep}(1e-3,3), \text{guess}=c(1,1,1))
\]

estimated parameters
0.04041047 1.298524 0.1290066

Testing $H_0$ against $H_1$

$H_0$: 0.0403 0.8922 0.1279
$H_1$: 0.04041047 1.298524 0.1290066

Divergence statistic: 8.7511 (p-value=0.24091)
Likelihood ratio test statistic: 6.883 (p-value=0.075723)
Now we run the test for $H_0 = H_1$, should not reject

```r
sdeDiv(X=X, theta0 = thetal, 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
Extensions

The multidimensional extension is also available and it is based entirely on the local gaussian approximation.
The multidimensional extension is also available and it is based entirely on the local Gaussian approximation.

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

\[ S_n(\phi) = \sup_x K_\phi(\hat{F}_n(x), x) \quad \text{and} \quad T_n(\phi) = \int_0^1 K_\phi(\hat{F}_n(x), x) \, dx \]

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.
Extensions

The multidimensional extension is also available and it is based entirely on the local gaussian approximation. In the i.i.d. case Jager and Wellner (2007) considered the statistics

\[ S_n(\phi) = \sup_x K_\phi(\hat{F}_n(x), x) \quad \text{and} \quad T_n(\phi) = \int_0^1 K_\phi(\hat{F}_n(x), x) \, dx \]

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.

Using Malliavin calculus, it may be possible to extend the results in Uchida and Yoshida (2004) to \( \phi \)-divergences.
Relevant references

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


