GENEVA SCHOOL OF ECONOMICS AND MANAGEMENT

UNIVERSITÉ DE GENÈVE

Saddlepoint techniques for the statistical analysis of time series

Davide La Vecchia

results from joint work with E. Ronchetti, A. Moor, C. Jiang & O. Scaillet

June-2024, Erasmus University

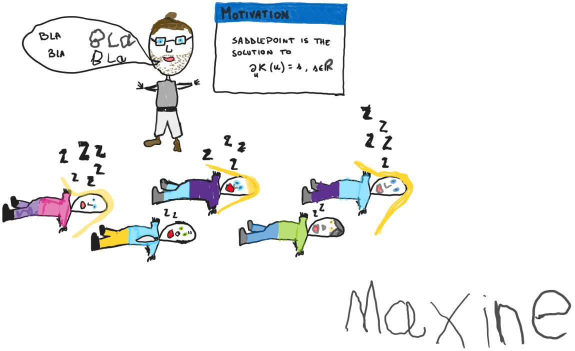
Before losing the audience due to a boring talk...

Before losing the audience due to a boring talk...

Maxine

Source: Maxine's drawing book, copyright fees paid by means of Swiss chocolate.

Before losing the audience due to a boring talk...



Source: Maxine's drawing book, copyright fees paid by means of Swiss chocolate.

... I state the aim of the talk ...

Three-fold aim of the talk

In the setting of time series with short- or long-memory, I:

Three-fold aim of the talk

In the setting of time series with short- or long-memory, I:

- Illustrate that first-order asymptotic theory suffers from finite sample distortions

Three-fold aim of the talk

In the setting of time series with short- or long-memory, I:

- Illustrate that first-order asymptotic theory suffers from finite sample distortions
- Develop saddlepoint techniques (for CDF/pdf approximation, p-values, and testing) which perform well in small samples and feature higher-order accuracy

Three-fold aim of the talk

In the setting of time series with short- or long-memory, I:

- Illustrate that first-order asymptotic theory suffers from finite sample distortions
- Develop saddlepoint techniques (for CDF/pdf approximation, p-values, and testing) which perform well in small samples and feature higher-order accuracy
- Illustrate numerically the performance of some saddlepoint techniques (testing)

Motivation from theoretical statistics

Typical statistical problem: For a given statistic/functional

$$T : \operatorname{dom} T \to \mathbb{R}^p, p \geq 1$$

(it can be an estimator $\hat{\theta}_n$), tail probabilities or quantiles at different levels are needed to carry out statistical inference (essentially, tests and confidence intervals).

Motivation from theoretical statistics

Typical statistical problem: For a given statistic/functional

$$T : \operatorname{dom} T \to \mathbb{R}^p, p \geq 1$$

(it can be an estimator $\hat{\theta}_n$), tail probabilities or quantiles at different levels are needed to carry out statistical inference (essentially, tests and confidence intervals).

Unless the (test) statistic/functional T or the estimator have a simple form (e.g. linear in the observations) and/or the underlying distribution of data has a particular form (e.g. normal), CDF/pdf or tail probabilities cannot be computed exactly.

Motivation from theoretical statistics

Typical statistical problem: For a given statistic/functional

$$T : \operatorname{dom} T \to \mathbb{R}^p, p \geq 1$$

(it can be an estimator $\hat{\theta}_n$), tail probabilities or quantiles at different levels are needed to carry out statistical inference (essentially, tests and confidence intervals).

Unless the (test) statistic/functional T or the estimator have a simple form (e.g. linear in the observations) and/or the underlying distribution of data has a particular form (e.g. normal), CDF/pdf or tail probabilities cannot be computed exactly.

$\Rightarrow$ we have to rely on approximations.

Motivation from financial econometrics

The same type of problem is relevant in other disciplines.

Motivation from financial econometrics

The same type of problem is relevant in other disciplines. For instance, in time series analysis

$$dY(t) = \mu(Y_t)dt + \sigma(Y_t)dW_t + J_t dN_t$$

where N_t is a Poisson process, J_t is the jump size, W_t is a BM.

Motivation from financial econometrics

The same type of problem is relevant in other disciplines. For instance, in time series analysis

$$dY(t) = \mu(Y_t)dt + \sigma(Y_t)dW_t + J_t dN_t$$

where N_t is a Poisson process, J_t is the jump size, W_t is a BM.

1. calculation of Value at Risk (VaR) or option prices: see e.g. Ait-Sahalia & Yu (2006, JoE), Glasserman & Kim (2009, JED&C), Rogers & Zane (1999, AoAP), Ait-Sahalia & Leaven (2023, wp). For VaR we need the CDF:

$$P(Y_{t+\Delta} \leq a_0 | Y_t = x)$$

Motivation from financial econometrics

The same type of problem is relevant in other disciplines. For instance, in time series analysis

$$dY(t) = \mu(Y_t)dt + \sigma(Y_t)dW_t + J_t dN_t$$

where N_t is a Poisson process, J_t is the jump size, W_t is a BM.

1. calculation of Value at Risk (VaR) or option prices: see e.g. Ait-Sahalia & Yu (2006, JoE), Glasserman & Kim (2009, JED&C), Rogers & Zane (1999, AoAP), Ait-Sahalia & Leaven (2023, wp). For VaR we need the CDF:

$$P(Y_{t+\Delta} \leq a_0 | Y_t = x)$$

2. by Fourier inversion, one gets the transition density (for $\Delta > 0$ and for $\tau \in \mathbb{R}$):

$$p(y|x,\Delta) = \frac{1}{2\pi i}\int_{\tau - i\infty}^{\tau + i\infty} \exp\left\{ K_{y|x}(\Delta, z; x) - zy \right\} dz$$

needed for inference on the model parameter; see e.g. Bibby et al. (Handbook of Fin. Econ., 2010), La Vecchia & Trojani, (JASA, 2012)

Motivation from financial econometrics

The same type of problem is relevant in other disciplines. For instance, in time series analysis

$$dY(t) = \mu(Y_t)dt + \sigma(Y_t)dW_t + J_t dN_t$$

where N_t is a Poisson process, J_t is the jump size, W_t is a BM.

1. calculation of Value at Risk (VaR) or option prices: see e.g. Ait-Sahalia & Yu (2006, JoE), Glasserman & Kim (2009, JED&C), Rogers & Zane (1999, AoAP), Ait-Sahalia & Leaven (2023, wp). For VaR we need the CDF:

$$P(Y_{t+\Delta} \leq a_0 | Y_t = x)$$

2. by Fourier inversion, one gets the transition density (for $\Delta > 0$ and for $\tau \in \mathbb{R}$):

$$p(y|x,\Delta) = \frac{1}{2\pi i}\int_{\tau - i\infty}^{\tau + i\infty} \exp\left\{ \boxed{K_{y|x}}(\Delta, z; x) - zy \right\} dz$$

needed for inference on the model parameter; see e.g. Bibby et al. (Handbook of Fin. Econ., 2010), La Vecchia & Trojani, (JASA, 2012)

3. distribution of estimators (e.g. Whittle's estimator, as in Phillips (1978, Biometrika), La Vecchia & Ronchetti (2019, JoE))

Motivation from financial econometrics

The same type of problem is relevant in other disciplines. For instance, in time series analysis

$$dY(t) = \mu(Y_t)dt + \sigma(Y_t)dW_t + J_t dN_t$$

where N_t is a Poisson process, J_t is the jump size, W_t is a BM.

1. calculation of Value at Risk (VaR) or option prices: see e.g. Ait-Sahalia & Yu (2006, JoE), Glasserman & Kim (2009, JED&C), Rogers & Zane (1999, AoAP), Ait-Sahalia & Leaven (2023, wp). For VaR we need the CDF:

$$P(Y_{t+\Delta} \leq a_0 | Y_t = x)$$

2. by Fourier inversion, one gets the transition density (for $\Delta > 0$ and for $\tau \in \mathbb{R}$):

$$p(y|x, \Delta) = \frac{1}{2\pi i} \int_{\tau - i\infty}^{\tau + i\infty} \exp\left\{ \boxed{K_{y|x}}(\Delta, z; x) - zy \right\} dz$$

needed for inference on the model parameter; see e.g. Bibby et al. (Handbook of Fin. Econ., 2010), La Vecchia & Trojani, (JASA, 2012)

3. distribution of estimators (e.g. Whittle's estimator, as in Phillips (1978, Biometrika), La Vecchia & Ronchetti (2019, JoE))
4. distribution of RV estimators, see e.g. Zhang et al. (2011, JoE)

Motivation from financial econometrics

The same type of problem is relevant in other disciplines. For instance, in time series analysis

$$dY(t) = \mu(Y_t)dt + \sigma(Y_t)dW_t + J_t dN_t$$

where N_t is a Poisson process, J_t is the jump size, W_t is a BM.

1. calculation of Value at Risk (VaR) or option prices: see e.g. Ait-Sahalia & Yu (2006, JoE), Glasserman & Kim (2009, JED&C), Rogers & Zane (1999, AoAP), Ait-Sahalia & Leaven (2023, wp). For VaR we need the CDF:

$$P(Y_{t+\Delta} \leq a_0 | Y_t = x)$$

2. by Fourier inversion, one gets the transition density (for $\Delta > 0$ and for $\tau \in \mathbb{R}$):

$$p(y|x,\Delta) = \frac{1}{2\pi i}\int_{\tau-i\infty}^{\tau+i\infty} \exp\left\{ \boxed{K_{y|x}}(\Delta, z; x) - zy \right\} dz$$

   needed for inference on the model parameter; see e.g. Bibby et al. (Handbook of Fin. Econ., 2010), La Vecchia & Trojani, (JASA, 2012)

3. distribution of estimators (e.g. Whittle's estimator, as in Phillips (1978, Biometrika), La Vecchia & Ronchetti (2019, JoE))
4. distribution of RV estimators, see e.g. Zhang et al. (2011, JoE)
5. ...

Possible solutions

We can approximate CDF/pdf/ or tail probabilities via

- *Asymptotic theory*: use of Central Limit Theorem to get a Gaussian approximation in large samples

Possible solutions

We can approximate CDF/pdf/ or tail probabilities via

- *Asymptotic theory*: use of Central Limit Theorem to get a Gaussian approximation in large samples
- *Analytical techniques*: use of expansions (Edgeworth, saddlepoint) to get an approximation in small samples

Possible solutions

We can approximate CDF/pdf/ or tail probabilities via

- *Asymptotic theory*: use of Central Limit Theorem to get a Gaussian approximation in large samples
- *Analytical techniques*: use of expansions (Edgeworth, saddlepoint) to get an approximation in small samples
- *Resampling techniques*: use of resampling (bootstrap, subsampling) computer-aided methods to get an approximation in small samples

Possible solutions

We can approximate CDF/pdf/ or tail probabilities via

- *Asymptotic theory*: use of Central Limit Theorem to get a Gaussian approximation in large samples
- *Analytical techniques*: use of expansions (Edgeworth, saddlepoint) to get an approximation in small samples
- *Resampling techniques*: use of resampling (bootstrap, subsampling) computer-aided methods to get an approximation in small samples

Analytical and resampling techniques can achieve higher order refinements over the first order asymptotic theory

First part: i.i.d. setting

Analytical techniques: Edgeworth expansion

Edgeworth/Charlier series is obtained as follows:

1. Assume we are given:

Random Variable	X	Y
Distr.	F_X	G
Measures	μ	ν
Charact. fct.	$\chi(u)$	$\xi(u)$
Cumulants	β_r	γ_r

where the two Fourier transforms are

$$\chi(u) = \int e^{iux} dF_X(x), \qquad \xi(u) = \int e^{iux} dG(x)$$

and the cumulants are (by definition)

$$\beta_r = (-i)^r \frac{d^r}{du^r} \ln \chi(u) \Big|_{u=0}, \qquad \gamma_r = (-i)^r \frac{d^r}{du^r} \ln \xi(u) \Big|_{u=0}$$

Analytical techniques: Edgeworth expansion

2 By **Taylor expansion of the cumulant generating function (c.g.f.) about** $u = 0$:

$$\ln \frac{\chi(u)}{\xi(u)} = \ln \chi(u) - \ln \xi(u) \overset{\text{Taylor at } u=0}{=} \sum_{r=1}^{\infty} (\beta_r - \gamma_r) \frac{(iu)^r}{r!},$$

thus,

$$\chi(u) = \exp\left\{ \sum_{r=1}^{\infty} (\beta_r - \gamma_r) \frac{(iu)^r}{r!} \right\} \xi(u).$$

Analytical techniques: Edgeworth expansion

2 By **Taylor expansion of the cumulant generating function (c.g.f.) about** $u = 0$:

$$\ln \frac{\chi(u)}{\xi(u)} = \ln \chi(u) - \ln \xi(u) \overset{\text{Taylor at } u=0}{=} \sum_{r=1}^{\infty} (\beta_r - \gamma_r) \frac{(iu)^r}{r!},$$

thus,

$$\chi(u) = \exp\left\{ \sum_{r=1}^{\infty} (\beta_r - \gamma_r) \frac{(iu)^r}{r!} \right\} \xi(u).$$

3 By Fourier inversion (under suitable assumptions on G, see Wallace (Ann. Math. Stats, 1958)):

$$F_X(x) = \exp\left\{ \sum_{r=1}^{\infty} (\beta_r - \gamma_r) \frac{(-D)^r}{r!} \right\} G(x),$$

D denotes a differentiation operator (with respect to x and $e^D = \sum_{j=0}^{\infty} D^j/j!$): it is such that the first term of the expansion of F_X is G.

Analytical techniques: Edgeworth expansion

2 By **Taylor expansion of the cumulant generating function (c.g.f.) about** $u = 0$:

$$\ln \frac{\chi(u)}{\xi(u)} = \ln \chi(u) - \ln \xi(u) \overset{\text{Taylor at } u=0}{=} \sum_{r=1}^{\infty} (\beta_r - \gamma_r) \frac{(iu)^r}{r!},$$

thus,

$$\chi(u) = \exp\left\{ \sum_{r=1}^{\infty} (\beta_r - \gamma_r) \frac{(iu)^r}{r!} \right\} \xi(u).$$

3 By Fourier inversion (under suitable assumptions on G, see Wallace (Ann. Math. Stats, 1958)):

$$F_X(x) = \exp\left\{ \sum_{r=1}^{\infty} (\beta_r - \gamma_r) \frac{(-D)^r}{r!} \right\} G(x),$$

D denotes a differentiation operator (with respect to x and $e^D = \sum_{j=0}^{\infty} D^j/j!$): it is such that the first term of the expansion of F_X is G.

4 Differentiation yields an expansion for the density of X.

Motivation form theoretical statistics

Example (Sample mean)

Let $X \sim \mu$ absolutely continuous w.r.t. the Lebesgue measure and having density f_X.

We are given a random sample $\boldsymbol{X} = (X_1, ..., X_n)$ of i.i.d. copies of X, whose m.g.f. and c.g.f. exist and: $E_\mu[X] = 0$, $V_\mu(X) = \sigma^2 < \infty$.

Motivation form theoretical statistics

Example (Sample mean)

Let $X \sim \mu$ absolutely continuous w.r.t. the Lebesgue measure and having density f_X.

We are given a random sample $\boldsymbol{X} = (X_1, ..., X_n)$ of i.i.d. copies of X, whose m.g.f. and c.g.f. exist and: $E_\mu[X] = 0$, $V_\mu(X) = \sigma^2 < \infty$.

The standardized mean (statistic, $T(\boldsymbol{X})$) has expression:

$$\sqrt{n}\frac{\bar{X}_n}{\sigma} = \sum_{i=1}^{n} \frac{X_i}{\sqrt{n}\sigma}.$$

Motivation form theoretical statistics

Example (Sample mean)

Let $X \sim \mu$ absolutely continuous w.r.t. the Lebesgue measure and having density f_X.

We are given a random sample $\boldsymbol{X} = (X_1, ..., X_n)$ of i.i.d. copies of X, whose m.g.f. and c.g.f. exist and: $E_\mu[X] = 0$, $V_\mu(X) = \sigma^2 < \infty$.

The standardized mean (statistic, $T(\boldsymbol{X})$) has expression:

$$\sqrt{n}\frac{\bar{X}_n}{\sigma} = \sum_{i=1}^{n} \frac{X_i}{\sqrt{n}\sigma}.$$

We make use of Step 1-3 to approximate the density f_n of the standardized mean via Edgeworth expansion...

Motivation form theoretical statistics

Example (Sample mean)

Let $X \sim \mu$ absolutely continuous w.r.t. the Lebesgue measure and having density f_X.

We are given a random sample $\boldsymbol{X} = (X_1, \ldots, X_n)$ of i.i.d. copies of X, whose m.g.f. and c.

T

W
E

Motivation form theoretical statistics

Example (cont'd)

... using as $G(x)$ the standard normal, we obtain an expansion of f_n in **powers of $n^{-1/2}$**, where the leading term is the standard normal density and higher order terms correct for skewness, kurtosis:

$$g_{\text{Edg}}(s) = \phi(s)$$

Motivation form theoretical statistics

Example (cont'd)

... using as $G(x)$ the standard normal, we obtain an expansion of f_n in **powers of $\boldsymbol{n^{-1/2}}$**, where the leading term is the standard normal density and higher order terms correct for skewness, kurtosis:

$$g_{\mathrm{Edg}}(s) = \phi(s) + \boldsymbol{n^{-1/2}} \frac{\lambda_3}{6}(s^3 - 3s)\phi(s)$$

Motivation form theoretical statistics

Example (cont'd)

... using as $G(x)$ the standard normal, we obtain an expansion of f_n in **powers of $\boldsymbol{n^{-1/2}}$**, where the leading term is the standard normal density and higher order terms correct for skewness, kurtosis:

$$
\begin{aligned}
g_{\mathrm{Edg}}(s) &= \phi(s) + \boldsymbol{n^{-1/2}} \frac{\lambda_3}{6}(s^3 - 3s)\phi(s) \\
&+ \boldsymbol{n^{-1}} \left[\frac{\lambda_4}{24}(s^4 - 6s^2 + 3) + \frac{\lambda_3^2}{72}(s^6 - 15s^4 + 45s^2 + 15) \right] \phi(s),
\end{aligned}
$$

with λ_3 and λ_4 being the standardized cumulants of X of order three and four, while ϕ is the pdf of a standard normal.

Motivation form theoretical statistics

Example (cont'd)

... using as $G(x)$ the standard normal, we obtain an expansion of f_n in **powers of $\boldsymbol{n^{-1/2}}$**, where the leading term is the standard normal density and higher order terms correct for skewness, kurtosis:

$$\begin{aligned} g_{\text{Edg}}(s) &= \phi(s) + \boldsymbol{n^{-1/2}} \frac{\lambda_3}{6}(s^3 - 3s)\phi(s) \\ &+ \boldsymbol{n^{-1}} \left[\frac{\lambda_4}{24}(s^4 - 6s^2 + 3) + \frac{\lambda_3^2}{72}(s^6 - 15s^4 + 45s^2 + 15) \right] \phi(s), \end{aligned}$$

with λ_3 and λ_4 being the standardized cumulants of X of order three and four, while ϕ is the pdf of a standard normal.

Remark

By construction, Edgeworth expansions provide in general a good approximation in the center of the density, **BUT**

Motivation form theoretical statistics

Example (cont'd)

... using as $G(x)$ the standard normal, we obtain an expansion of f_n in **powers of $\boldsymbol{n^{-1/2}}$**, where the leading term is the standard normal density and higher order terms correct for skewness, kurtosis:

$$
\begin{aligned}
g_{\text{Edg}}(s) &= \phi(s) + \boldsymbol{n^{-1/2}} \frac{\lambda_3}{6}(s^3 - 3s)\phi(s) \\
&+ \boldsymbol{n^{-1}} \left[\frac{\lambda_4}{24}(s^4 - 6s^2 + 3) + \frac{\lambda_3^2}{72}(s^6 - 15s^4 + 45s^2 + 15) \right] \phi(s),
\end{aligned}
$$

with λ_3 and λ_4 being the standardized cumulants of X of order three and four, while ϕ is the pdf of a standard normal.

Remark

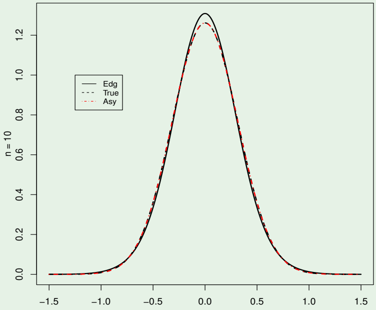
By construction, Edgeworth expansions provide in general a good approximation in the center of the density, **BUT** *they can be* **inaccurate in the tails**, *where they can even become* **negative**.

Motivation form theoretical statistics

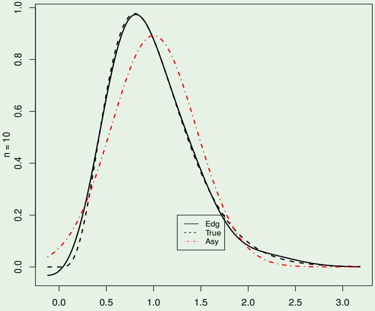
Example (cont'd)

To have a graphical illustration, for Asy and Edg, consider $\bar{X}_n$ (use Jacobian formula on the dens approx for the standardized mean) for $n = 10, 50, 250$, for $X_i \sim \mathcal{N}(0,1)$ and $X_i \sim \exp(1)$

$\mathcal{N}(0,1)$ $(n = 10)$



$\exp(1)$ $(n = 10)$

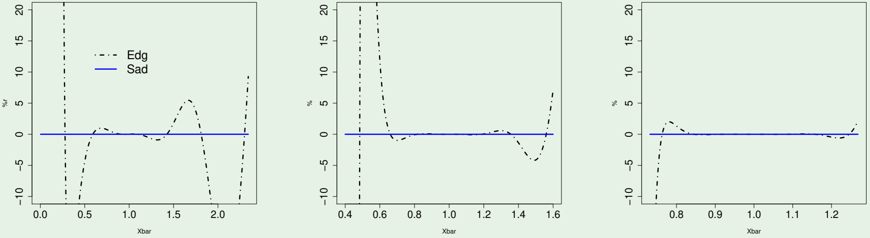


Motivation form theoretical statistics

Example (cont'd)

for the exponential case, rel. err. $= 100 \cdot$ (true -approx)/true

$n = 10$ $n = 50$ $n = 250$



Any other higher oder technique to cope with these issues? saddlepoint approx...

Motivation form theoretical statistics

Example (cont'd)

In this example about $\bar{X}_n$, we know the c.g.f. and the saddlepoint density approx $g_n(s)$ is (Daniels (1954)):

$$g_n(s) = \left[\frac{n}{2\pi \mathcal{K}'' \{v(s)\}}\right]^{1/2} \exp\left(n\left[\mathcal{K}\{v(s)\} - v(s)s\right]\right) \tag{1}$$

and $v(s)$ (saddlepoint) is the solution to

$$\mathcal{K}'(v) - s = 0,$$

Motivation form theoretical statistics

Example (cont'd)

In this example about $\bar{X}_n$, we know the c.g.f. and the saddlepoint density approx $g_n(s)$ is (Daniels (1954)):

$$g_n(s) = \left[\frac{n}{2\pi \mathcal{K}''\{v(s)\}}\right]^{1/2} \exp\left(n\left[\mathcal{K}\{v(s)\} - v(s)s\right]\right) \tag{1}$$

and $v(s)$ (saddlepoint) is the solution to

$$\mathcal{K}'(v) - s = 0,$$

namely, we look for $v(s)$ such that X has expected value equal to s.

Motivation form theoretical statistics

Example (cont'd)

- $g_n(s)$ is a "Gaussian-type" integral with both mean and variance that depends on s: it is a density-like object that cannot take on negative values ($\neq$ Edg):

Motivation form theoretical statistics

Example (cont'd)

- $g_n(s)$ is a "Gaussian-type" integral with both mean and variance that depends on s: it is a density-like object that cannot take on negative values ($\neq$ Edg):
  E.g. for i.i.d. standard Gaussian rvs:

$$\mathcal{K}(v) = \frac{v^2}{2}, \quad \mathcal{K}'(v) = v \text{ and } \mathcal{K}''(v) = 1,$$

  the saddlepoint is defined by $\mathcal{K}'(v) - s = 0$, thus $v(s) = s$ and

Motivation form theoretical statistics

Example (cont'd)

- $g_n(s)$ is a "Gaussian-type" integral with both mean and variance that depends on s: it is a density-like object that cannot take on negative values ($\neq$ Edg):
  E.g. for i.i.d. standard Gaussian rvs:

$$\mathcal{K}(\upsilon) = \frac{\nu^2}{2}, \quad \mathcal{K}'(\upsilon) = \upsilon \text{ and } \mathcal{K}''(\upsilon) = 1,$$

  the saddlepoint is defined by $\mathcal{K}'(\upsilon) - s = 0$, thus $\upsilon(s) = s$ and

- for $\bar{X}_n$, $g_n(s) = \left[\frac{n}{2\pi}\right]^{1/2} e^{-\frac{ns^2}{2}}$

Motivation form theoretical statistics

Example (cont'd)

- $g_n(s)$ is a "Gaussian-type" integral with both mean and variance that depends on s: it is a density-like object that cannot take on negative values ($\neq$ Edg):
  E.g. for i.i.d. standard Gaussian rvs:

$$\mathcal{K}(v) = \frac{v^2}{2}, \quad \mathcal{K}'(v) = v \text{ and } \mathcal{K}''(v) = 1,$$

the saddlepoint is defined by $\mathcal{K}'(v) - s = 0$, thus $v(s) = s$ and

- for $\bar{X}_n$, $g_n(s) = \left[\frac{n}{2\pi}\right]^{1/2} e^{-\frac{ns^2}{2}}$ pdf of $\mathcal{N}\left(0, \frac{1}{n}\right)$ it's exact!!

Motivation form theoretical statistics

Example (cont'd)

- $g_n(s)$ is a "Gaussian-type" integral with both mean and variance that depends on s: it is a density-like object that cannot take on negative values ($\neq$ Edg):
  E.g. for i.i.d. standard Gaussian rvs:

$$\boxed{\mathcal{K}}(\upsilon) = \frac{\nu^2}{2}, \quad \boxed{\mathcal{K}'}(\upsilon) = \upsilon \text{ and } \boxed{\mathcal{K}''}(\upsilon) = 1,$$

  the saddlepoint is defined by $\boxed{\mathcal{K}'}(\upsilon) - s = 0$, thus $\upsilon(s) = s$ and

- for $\bar{X}_n$, $g_n(s) = \left[\frac{n}{2\pi}\right]^{1/2} e^{-\frac{ns^2}{2}}$ pdf of $\mathcal{N}\left(0, \frac{1}{n}\right)$ it's exact!!
- for $\sqrt{n}\bar{X}_n$, (by Jacobian formula) $g_n(s) = \left[\frac{1}{2\pi}\right]^{1/2} e^{-\frac{s^2}{2}}$

Motivation form theoretical statistics

Example (cont'd)

- $g_n(s)$ is a "Gaussian-type" integral with both mean and variance that depends on s: it is a density-like object that cannot take on negative values ($\neq$ Edg):
  E.g. for i.i.d. standard Gaussian rvs:

$$\mathcal{K}(v) = \frac{\nu^2}{2}, \quad \mathcal{K}'(v) = v \text{ and } \mathcal{K}''(v) = 1,$$

  the saddlepoint is defined by $\mathcal{K}'(v) - s = 0$, thus $v(s) = s$ and

- for $\bar{X}_n$, $g_n(s) = \left[\frac{n}{2\pi}\right]^{1/2} e^{-\frac{ns^2}{2}}$ pdf of $\mathcal{N}\left(0, \frac{1}{n}\right)$ it's exact!!
- for $\sqrt{n}\bar{X}_n$, (by Jacobian formula) $g_n(s) = \left[\frac{1}{2\pi}\right]^{1/2} e^{-\frac{s^2}{2}}$ pdf of $\mathcal{N}(0, 1)$ it's exact!!

Motivation form theoretical statistics

Example (cont'd)

- The saddlepoint density approximation g_n features relative error of order $O(n^{-1})$ over the whole $\mathbb{R}$

$$f_n(s) = g_n(s)\left\{1 + O\left(n^{-1}\right)\right\} \tag{2}$$

Motivation form theoretical statistics

Example (cont'd)

- The saddlepoint density approximation g_n features relative error of order $O(n^{-1})$ over the whole $\mathbb{R}$

$$f_n(s) = g_n(s)\left\{1 + O\left(n^{-1}\right)\right\} \tag{2}$$

- The density g_n is obtained by approximating the Fourier inversion of M^n, which yields f_n:

$$\begin{aligned} f_n(s) &= \frac{n}{2\pi}\int_{-\infty}^{\infty} e^{-i\upsilon n s} M^n(i\upsilon)d\upsilon \overset{(z=i\upsilon)}{=} \frac{n}{2\pi i}\int_{\mathcal{I}} e^{-nzs} M^n(z)dz \\ &= \frac{n}{2\pi i}\int_{\tau-i\infty}^{\tau+i\infty} e^{n(\mathcal{K}(z)-zs)}dz, \ \tau \in \mathbb{R}, \end{aligned}$$

which may be obtained using a Taylor expansion of $(\mathcal{K}(z) - zs)$ about $\upsilon(s)$.

Motivation form theoretical statistics

The sadd approx g_n is obtained via the method of the steepest descent: this is a general technique in complex analysis applied to compute asymptotic expansions of integrals

$$\int_{\mathcal{P}} e^{v\, w(z)} \xi(z) dz,$$

with $v \in \mathbb{R}^+$ is large, ξ and w being analytic functions of $z \in \mathbb{C}$.

Motivation form theoretical statistics

The sadd approx g_n is obtained via the method of the steepest descent: this is a general technique in complex analysis applied to compute asymptotic expansions of integrals

$$\int_{\mathcal{P}} e^{v\, \boxed{w}(z)} \xi(z) dz,$$

with $v \in \mathbb{R}^+$ is large, ξ and w being analytic functions of $z \in \mathbb{C}$.

Idea

Deform the path of integration (Cauchy's theorem) so that the new path of integration passes through the so-called saddlepoint, namely the zero of the derivative $\boxed{w'}(z)$. Then, we approximate the resulting integral using a series expansion (Watson's lemma). See Daniels (AoMS, 1954).

Loosely speaking, one dos a "Laplace-type approx" on $\mathbb{C}$. Jump to Laplace

Motivation form theoretical statistics

Alternative: derive g_n via convex analysis.

Motivation form theoretical statistics

Alternative: derive g_n via convex analysis.

Idea (Sadd from Edg)

We rely on the method of the conjugate density or tilted Edgeworth:

Motivation form theoretical statistics

Alternative: derive g_n via convex analysis.

Idea (Sadd from Edg)

We rely on the method of the conjugate density or tilted Edgeworth:

- **by means of $\upsilon(s)$, recenter/Esscher tilt** *the density of X: we embed the original density f_X into an exponential family, and then define the (conjugate) density h_s such that it centers at s the density of the rv ($f_X \mapsto h_s$ via $\upsilon(s)$)*

Motivation form theoretical statistics

Alternative: derive g_n via convex analysis.

Idea (Sadd from Edg)

We rely on the method of the conjugate density or tilted Edgeworth:

- **by means of $\upsilon(s)$, recenter/Esscher tilt** *the density of X: we embed the original density f_X into an exponential family, and then define the (conjugate) density h_s such that it centers at s the density of the rv ($f_X \mapsto h_s$ via $\upsilon(s)$)*
- *compute a low-order* **Edgeworth expansion** *for the tilted density (centered at s, so it works well!) to obtain $g_n(s)$*

Motivation form theoretical statistics

Alternative: derive g_n via convex analysis.

Idea (Sadd from Edg)

We rely on the method of the conjugate density or tilted Edgeworth:

- **by means of** $\upsilon(s)$**, recenter/Esscher tilt** *the density of* X*: we embed the original density* f_X *into an exponential family, and then define the (conjugate) density* h_s *such that it centers at s the density of the rv* ($f_X \mapsto h_s$ *via* $\upsilon(s)$)
- *compute a low-order* **Edgeworth expansion** *for the tilted density (centered at s, so it works well!) to obtain* $g_n(s)$
- **repeat** *this procedure* **for every** $s \in \mathbb{R}$

Motivation form theoretical statistics

Alternative: derive g_n via convex analysis.

Idea (Sadd from Edg)

We rely on the method of the conjugate density or tilted Edgeworth:

- **by means of** $\upsilon(s)$**, recenter/Esscher tilt** *the density of* X*: we embed the original density* f_X *into an exponential family, and then define the (conjugate) density* h_s *such that it centers at* s *the density of the rv* ($f_X \mapsto h_s$ *via* $\upsilon(s)$)
- *compute a low-order* **Edgeworth expansion** *for the tilted density (centered at* s*, so it works well!) to obtain* $g_n(s)$
- **repeat** *this procedure* **for every** $s \in \mathbb{R}$

$\Rightarrow$ *saddlepoint density approximation is a sequence of low-order local approximations; see Easton & Ronchetti (1986), JASA, Wang (1992) and La Vecchia et al. (2023).*

Second part: time series setting

Motivations related to dependent data analysis

Many macroeconomic or environmental time series display a persistent time trend and contain only a few observations recorded at annual frequency. Focusing on econometrics, much controversy has revolved around the suitability of ARIMA models; see the seminal paper of Nelson and Plosser (1982) and Gil-Alana and Robinson (1997) for a review of the literature.

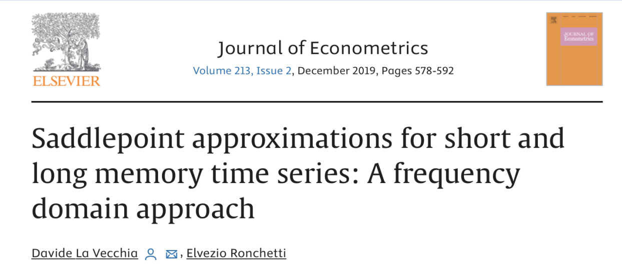
Motivations related to dependent data analysis

Many macroeconomic or environmental time series display a persistent time trend and contain only a few observations recorded at annual frequency. Focusing on econometrics, much controversy has revolved around the suitability of ARIMA models; see the seminal paper of Nelson and Plosser (1982) and Gil-Alana and Robinson (1997) for a review of the literature.

Within this setting, to model the slow decay of the autocorrelation function displayed by many macroeconomic time series, the use of (Gaussian) FARIMA models and first order Gaussian asymptotic theory (Wald-type test statistics) is routinely applied for confidence intervals and testing statistical hypotheses.

Motivations related to dependent data analysis

Many macroeconomic or environmental time series display a persistent time trend and contain only a few observations recorded at annual frequency. Focusing on econometrics, much controversy has revolved around the suitability of ARIMA models; see the seminal paper of Nelson and Plosser (1982) and Gil-Alana and Robinson (1997) for a review of the literature.



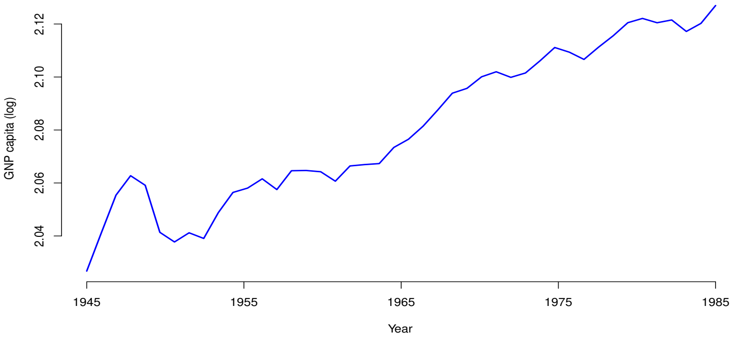
Journal of Econometrics
Volume 213, Issue 2, December 2019, Pages 578-592

Saddlepoint approximations for short and long memory time series: A frequency domain approach

Davide La Vecchia, Elvezio Ronchetti

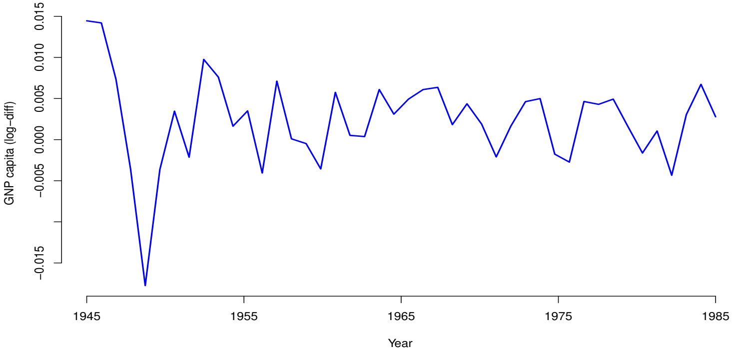
Motivations related to dependent data analysis

Focus on the extended Nelson and Plosser data set: plot log-GNP per capita (other time series available in the JoE paper)



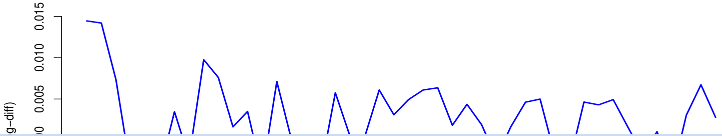
Motivations related to dependent data analysis

Focus on the extended Nelson and Plosser data set: plot log-diff GNP per capita (other time series available in the JoE paper)



Motivations related to dependent data analysis

Focus on the extended Nelson and Plosser data set: plot log-diff GNP per capita (other time series available in the JoE paper)



Remark

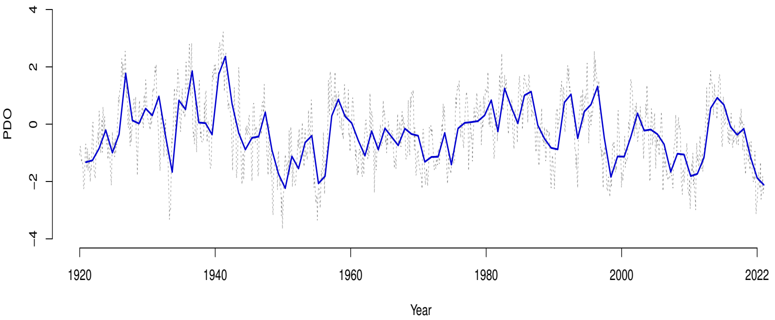
In the literature one is typically testing for the presence of long memory: ARFIMA models and

$$\mathcal{H}_0: \; d = 0 \quad vs \quad \mathcal{H}_1: \; d > 0$$

we resort on an M-estimator (Whittle), which is asymptotically χ^2 Wald-type test statistics (W_n) are applied when $n = 44$. Is this a sensible procedure? Is the asymptotics suffering from size distortion due to the small sample size?

Motivations related to dependent data analysis

The Pacific Decadal Oscillation (PDO) index measures the climatological situation of the Southern hemisphere: its extremes correspond to episodes of abnormal weather conditions.



Motivations related to dependent data analysis

The Pacific Decadal Oscillation (PDO) index measures the climatological situation of the Southern hemisphere: its extremes correspond to episodes of abnormal weather conditions.

Remark

Whiting et al. (2003) model the time series by an $ARFIMA(0, d, 0)$. *Data analysis and inference is conducted using* **annual data**, *from 1920 to 2022, so* $n = 122$, *relying on M-estimator (Whittle), which yields Wald-type statistic* W_n *from first order asymptotic theory to test*

$$\mathcal{H}_0 : \; d = 0 \quad vs \quad \mathcal{H}_1 : \; d > 0.$$

Motivations related to dependent data analysis

Example (ARFIMA synthetic data)

Let $\{Y_t, \ t \in \mathbb{Z}\}$ be an ARFIMA(p, d, q), having dynamics

$$\theta(L)(1-L)^d Y_t = \phi(L)\epsilon_t, \tag{3}$$

where $\forall t$, the $\{\epsilon_t\}$ are i.i.d. with zero mean and known $\sigma_\epsilon^2 = 1$.

Motivations related to dependent data analysis

Example (ARFIMA synthetic data)

Let $\{Y_t,\ t \in \mathbb{Z}\}$ be an ARFIMA(p, d, q), having dynamics

$$\theta(L)(1-L)^d Y_t = \phi(L)\epsilon_t, \tag{3}$$

where $\forall t$, the $\{\epsilon_t\}$ are i.i.d. with zero mean and known $\sigma_\epsilon^2 = 1$.

- We focus on a Gaussian ARFIMA of order $p = 2$, $d = 0$ and $q = 0$ process, with AR coefficients $\theta_1 = 0.1$ and $\theta_2 = 0.2$.

Motivations related to dependent data analysis

Example (ARFIMA synthetic data)

Let $\{Y_t,\ t \in \mathbb{Z}\}$ be an ARFIMA(p, d, q), having dynamics

$$\theta(L)(1-L)^d Y_t = \phi(L)\epsilon_t, \tag{3}$$

where $\forall t$, the $\{\epsilon_t\}$ are i.i.d. with zero mean and known $\sigma_\epsilon^2 = 1$.

- We focus on a Gaussian ARFIMA of order $p = 2$, $d = 0$ and $q = 0$ process, with AR coefficients $\theta_1 = 0.1$ and $\theta_2 = 0.2$.
- We consider increasing values of the sample size $n = 250, 2500, 5000$.

Motivations related to dependent data analysis

Example (ARFIMA synthetic data)

Let $\{Y_t,\ t \in \mathbb{Z}\}$ be an ARFIMA(p, d, q), having dynamics

$$\theta(L)(1-L)^d Y_t = \phi(L)\epsilon_t, \tag{3}$$

where $\forall t$, the $\{\epsilon_t\}$ are i.i.d. with zero mean and known $\sigma_\epsilon^2 = 1$.

- We focus on a Gaussian ARFIMA of order $p = 2$, $d = 0$ and $q = 0$ process, with AR coefficients $\theta_1 = 0.1$ and $\theta_2 = 0.2$.
- We consider increasing values of the sample size $n = 250, 2500, 5000$.
- We estimate θ via the routinely applied Whittle's M-estimator, as implemented in the routine `WhittleEst` available in the R package `longmemo`.

Motivations related to dependent data analysis

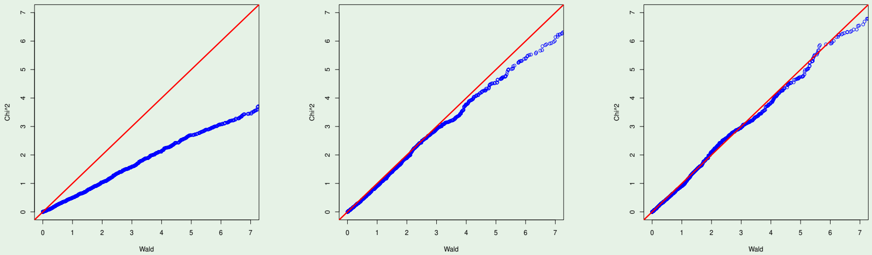
Example (cont'd)

The goal of our inference is to test

$$\mathcal{H}_0 : d = 0 \text{ vs. } \mathcal{H}_1 : d > 0,$$

and we resort on the Wald test statistic W_n for Whittle's estimator, as available in the statistical software, comparing χ^2 quantiles to the true (as obtained by MC simulation).

$n = 250$ $\quad$ $n = 2500$ $\quad$ $n = 5000$



Motivations related to dependent data analysis

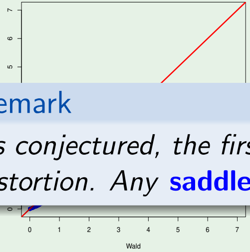
Example (cont'd)

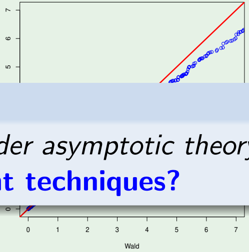
The goal of our inference is to test

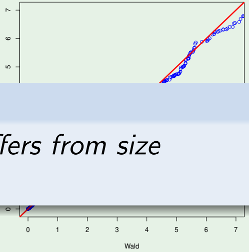
$$\mathcal{H}_0 : d = 0 \text{ vs. } \mathcal{H}_1 : d > 0,$$

and we resort on the Wald test statistic W_n for Whittle's estimator, as available in the statistical software, comparing χ^2 quantiles to the true (as obtained by MC simulation).

$n = 250$ $n = 2500$ $n = 5000$







Remark

As conjectured, the first order asymptotic theory suffers from size distortion. Any **saddlepoint techniques?**

Motivations related to dependent data analysis

Another example comes from the literature on Spatial Autoregressive processes (random fields):

Menu

- Literature: a bird's-eye view

Menu

- Literature: a bird's-eye view
- Time series
    - Some elements of spectral analysis
    - Setting: SRD & LRD

Menu

- Literature: a bird's-eye view
- Time series
  - Some elements of spectral analysis
  - Setting: SRD & LRD
  - Saddlepoint techniques

Menu

- Literature: a bird's-eye view
- Time series
  - Some elements of spectral analysis
  - Setting: SRD & LRD
  - Saddlepoint techniques
    - exponential-based (density approx and test in the presence of nuisance parameter)

Menu

- Literature: a bird's-eye view
- Time series
  - Some elements of spectral analysis
  - Setting: SRD & LRD
  - Saddlepoint techniques
    - exponential-based (density approx and test in the presence of nuisance parameter)
    - empirical version (density approx and test in the presence of nuisance parameter, connection to EL & ET)

Menu

- Literature: a bird's-eye view
- Time series
  - Some elements of spectral analysis
  - Setting: SRD & LRD
  - Saddlepoint techniques
    - exponential-based (density approx and test in the presence of nuisance parameter)
    - empirical version (density approx and test in the presence of nuisance parameter, connection to EL & ET)
  - Monte Carlo results

Literature: a bird's-eye view

(i) Most of the results on **saddlepoint techniques** are available for the **iid setting:** see Field & Ronchetti (1990), Jensen (1995), Kolassa (2006), Butler (2007), or Brazzale et al. (2007) for book-length presentation.

Literature: a bird's-eye view

(i) Most of the results on **saddlepoint techniques** are available for the **iid setting:** see Field & Ronchetti (1990), Jensen (1995), Kolassa (2006), Butler (2007), or Brazzale et al. (2007) for book-length presentation.

(ii) In time series, some results for **saddlepoint density and tail area approximations for M-estimators** in the noncircular **Gaussian AR(1) process**: see Daniels (Annals, 1956), Phillips (Biometrika, 1978), Cox & Solomon (Biometrika, 1988), Wang (Biometrika, 1992), Butler & Paolella (Bernoulli, 1998), Pereira et al. (Stats & Prob Letters, 2008), Lozada-Can & Davison (American Statistician, 2010),and Field & Robinson (Annals, 2013); for a book-length discussion see Taniguchi & Kakitzawa (2001).

Literature: a bird's-eye view

(i) Most of the results on **saddlepoint techniques** are available for the **iid setting:** see Field & Ronchetti (1990), Jensen (1995), Kolassa (2006), Butler (2007), or Brazzale et al. (2007) for book-length presentation.

(ii) In time series, some results for **saddlepoint density and tail area approximations for *M*-estimators** in the noncircular **Gaussian AR(1) process**: see Daniels (Annals, 1956), Phillips (Biometrika, 1978), Cox & Solomon (Biometrika, 1988), Wang (Biometrika, 1992), Butler & Paolella (Bernoulli, 1998), Pereira et al. (Stats & Prob Letters, 2008), Lozada-Can & Davison (American Statistician, 2010),and Field & Robinson (Annals, 2013); for a book-length discussion see Taniguchi & Kakitzawa (2001).

(iii) **Higher order techniques in frequency domain (spectral analysis) for time series** are available: see Taniguchi (JMA, 1987, Edgeworth for Whittle under SRD), Franke & Härdle (Annals, 1992, FDB), Dahlhaus & Janas (Annals, 1996, FDB), Andrews & Lieberman (Econometric Theory, 2005, Edgeworth for Whittle under LRD).

Some elements of spectral analysis

To start with, let me recall the autocovariance function

$$\gamma_Y(h) = cov(Y_{t+h}, Y_t) = E[(Y_{t+h} - \mu)(Y_t - \mu)]$$

for $h \in \mathbb{Z}$ and with $E(Y_t) = \mu$, $\forall t$.

Some elements of spectral analysis

To start with, let me recall the autocovariance function

$$\gamma_Y(h) = cov(Y_{t+h}, Y_t) = E[(Y_{t+h} - \mu)(Y_t - \mu)]$$

for $h \in \mathbb{Z}$ and with $E(Y_t) = \mu$, $\forall t$.
Under suitable assumptions, we have (for $i \in \mathbb{C}$)

$$\gamma_Y(h) = \int_{-1/2}^{1/2} \exp\{2\pi i \lambda h\} f(\lambda) d\lambda, \quad h = 0, \pm 1, \pm 2 ...$$

as the inverse Fourier transform of the spectral density $f(\cdot)$:

$$f(\lambda) = \sum_{h=-\infty}^{\infty} \gamma_Y(h) \exp\{-i2\pi\lambda h\}, \quad -1/2 \le \lambda \le 1/2.$$

Some elements of spectral analysis

Definition

Given time series data $Y_1, \ldots, Y_n$, the discrete Fourier transform (DFT) is

$$d(\lambda_j) = n^{-1/2} \sum_{t=1}^{n} Y_t \exp\{-2\pi i \lambda_j t\},$$

for $j = 0, 1, \ldots, n-1$, where the frequencies $\lambda_j = j/n$ are called Fourier or fundamental frequencies.

Some elements of spectral analysis

Definition

Given time series data $Y_1, \ldots, Y_n$, the discrete Fourier transform (DFT) is

$$d(\lambda_j) = n^{-1/2} \sum_{t=1}^{n} Y_t \exp\{-2\pi i \lambda_j t\},$$

for $j = 0, 1, \ldots, n-1$, where the frequencies $\lambda_j = j/n$ are called Fourier or fundamental frequencies.

The periodogram at λ_j is $I(\lambda_j) = |d(\lambda_j)|^2$, and we have that

$$I(\lambda_j) = \sum_{h=-(n-1)}^{n-1} \hat{\gamma}_Y(h) \exp\{-2i\pi\lambda_j h\},$$

where $\hat{\gamma}_Y(h)$ is the empirical (auto)covariance.

Some elements of spectral analysis

Property. The periodogram ordinates are such that, as $n \to \infty$:

$$I(\lambda) \stackrel{\mathcal{D}}{\to} i.d.\ \xi f(\lambda), \quad \xi \sim \text{exp}^*(1) \tag{4}$$

Remark

The asymptotic iid-ness of the standardized periodogram ordinates allows to transform problems for dependent data into problems for iid data.

Some elements of spectral analysis

The Property allows to derive a frequency domain likelihood and parameter estimation is obtained maximazing this likelihood.

Some elements of spectral analysis

The Property allows to derive a frequency domain likelihood and parameter estimation is obtained maximazing this likelihood.

This idea goes back to Whittle (1951): if there is a parametric model for $f(\lambda, \theta)$, then:

$$\mathcal{L}_W(\theta) = \frac{1}{2\pi}\left[\int_{-\pi}^{\pi} \ln f(\lambda,\theta) d\lambda + \int_{-\pi}^{\pi} \frac{I(\lambda)}{f(\lambda,\theta)} d\lambda\right], \tag{5}$$

which is obtained making use of Property (λ is in radians, from now on).

Some elements of spectral analysis

The Property allows to derive a frequency domain likelihood and parameter estimation is obtained maximazing this likelihood.

This idea goes back to Whittle (1951): if there is a parametric model for $f(\lambda,\theta)$, then:

$$\mathcal{L}_W(\theta) = \frac{1}{2\pi}\left[\int_{-\pi}^{\pi} \ln f(\lambda,\theta)d\lambda + \int_{-\pi}^{\pi} \frac{I(\lambda)}{f(\lambda,\theta)}d\lambda\right], \tag{5}$$

which is obtained making use of Property (λ is in radians, from now on).

The optimization of $L_W(\theta)$ (the Riemann-discretized version of $\mathcal{L}_W$):

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

(or $\nabla_\theta L_W(\hat{\theta}_n) = 0$) defines an **M-estimator in the frequency domain, having estimating function** ψ_j. Then,

$$\mathcal{V}_n = \sqrt{n}(\hat{\theta}_n - \theta^0)$$

and we want an approximation to its density $f_{\hat{\theta}_n}$.

Incidentally....

At the other edge of the spectrum, the frequency domain approach can be suitably adapted and exploited for the statistical analysis of high-dimensional (large time, spatial, cross-sectional dimensions) random fields over a regular network:

arXiv > stat > arXiv:2312.02591

Search...
Help | Ac

Statistics > Methodology

[Submitted on 5 Dec 2023]

General Spatio-Temporal Factor Models for High-Dimensional Random Fields on a Lattice

Matteo Barigozzi, Davide La Vecchia, Hang Liu

Motivated by the need for analysing large spatio-temporal panel data, we introduce a novel dimensionality reduction methodology for n-dimensional random fields observed across a number S spatial locations and T time periods. We call it General Spatio-Temporal Factor Model (GSTFM). First, we provide the probabilistic and mathematical underpinning needed for the representation of a random field as the sum of two components: the common component (driven by a small number q of latent factors) and the idiosyncratic component (mildly cross-correlated). We show that the two components are identified as $n \to \infty$. Second, we propose an estimator of the common component and derive its statistical guarantees (consistency and rate of convergence) as $\min(n, S, T) \to \infty$. Third, we propose an information criterion to determine the number of factors. Estimation makes use of Fourier analysis in the frequency domain and thus we fully exploit the information on the spatio-temporal covariance structure of the whole panel. Synthetic data examples illustrate the applicability of GSTFM and its advantages over the extant generalized dynamic factor model that ignores the spatial correlations.

Setting: SRD and LRD

For the sake of exposition, suppose that $\{Y_t\}$ admits is a linear second order stationary process (see e.g. Hannan (1970), Ch. 2 or Beran et al. (2010), Ch. 2) with spectral density function

$$f(\lambda, \theta) = |\lambda|^{-2d} L(\lambda, \vartheta), \quad \lambda \in \Pi = (-\pi, \pi] \tag{6}$$

where $d \in [0, 0.5)$, $\vartheta \in \mathbb{R}^p$ with $p \geq 1$ and $\theta = (d, \vartheta)$.

Setting: SRD and LRD

For the sake of exposition, suppose that $\{Y_t\}$ admits is a linear second order stationary process (see e.g. Hannan (1970), Ch. 2 or Beran et al. (2010), Ch. 2) with spectral density function

$$f(\lambda, \theta) = |\lambda|^{-2d} L(\lambda, \vartheta), \quad \lambda \in \Pi = (-\pi, \pi] \tag{6}$$

where $d \in [0, 0.5)$, $\vartheta \in \mathbb{R}^p$ with $p \geq 1$ and $\theta = (d, \vartheta)$.

Definition

We classify the process $\{Y_t\}$ as short-range dependent (SRD) or long-range dependent (LRD)

- when $d = 0$ and the function $L(\cdot, \vartheta)$ is bounded with $L(0, \vartheta) \neq 0$, then the process $\{Y_t\}$ features SRD
- Otherwise, the process $\{Y_t\}$ features LRD—f has a pole at $\lambda = 0$.

Saddlepoint approximation (exponential-based)

First order asymptotic theory for Whittle's estimator implies

$$\mathcal{V}_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, V).$$

where V depends on the 2nd moment of ψ_j and the first moment of $\nabla_\theta \psi_j$.

Saddlepoint approximation (exponential-based)

First order asymptotic theory for Whittle's estimator implies

$$\mathcal{V}_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, V).$$

where V depends on the 2nd moment of ψ_j and the first moment of $\nabla_\theta \psi_j$.

To have a better density approximation to $f_{\hat{\theta}_n}$, we may derive the saddlepoint density approximation $g_{\hat{\theta}_n}$ treating the periodogram ordinates as independently and exponentially distributed r.v.'s: we use it to approximate the c.g.f. $\mathcal{K}_{\mathcal{V}_n}$ ...

Saddlepoint approximation (exponential-based)

First order asymptotic theory for Whittle's estimator implies

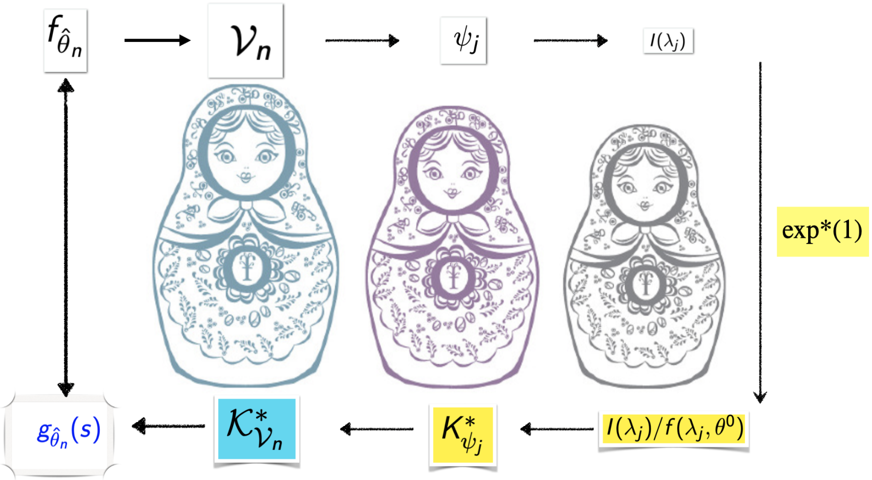
$$\mathcal{V}_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, V).$$

where V depends on the 2nd moment of ψ_j and the first moment of $\nabla_\theta \psi_j$.

To have a better density approximation to $f_{\hat{\theta}_n}$, we may derive the saddlepoint density approximation $g_{\hat{\theta}_n}$ treating the periodogram ordinates as independently and exponentially distributed r.v.'s: we use it to approximate the c.g.f. $\mathcal{K}_{\mathcal{V}_n}$ ...

... let's visualize the idea ...

Visualisation

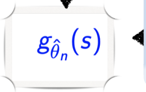


Visualisation

The saddlepoint approximation $g_{\hat{\theta}_n}$ can be **easily** derived treating the periodogram ordinates $\{I(\lambda)\}$ as independent rvs, exponentially distributed.

It features:

- **SRD**: relative of order $o(n^{-1/2})$
- **LRD**: relative error of order $O(n^{-1/2})$.



Saddlepoint approximation (exponential-based)

Specifically:

- Whittle's estimating function is

$$\psi_j\left(I(\lambda_j),\theta\right) = \left(\frac{I(\lambda_j)}{f(\lambda_j,\theta)} - 1\right)\nabla_\theta \ln f(\lambda_j,\theta),$$

Saddlepoint approximation (exponential-based)

Specifically:

- Whittle's estimating function is

$$\psi_j\left(I(\lambda_j),\theta\right) = \left(\frac{I(\lambda_j)}{f(\lambda_j,\theta)} - 1\right)\nabla_\theta \ln f(\lambda_j,\theta),$$

- for $m = \lfloor (n-1)/2 \rfloor$, Whittle's M-estimator $\hat{\theta}_n$ is the solution to

$$\sum_{j=1}^{m} \psi_j(I(\lambda_j),\hat{\theta}_n) = 0.$$

Saddlepoint approximation (exponential-based)

Specifically:

- Whittle's estimating function is

$$\psi_j\left(I(\lambda_j),\theta\right) = \left(\frac{I(\lambda_j)}{f(\lambda_j,\theta)} - 1\right)\nabla_\theta \ln f(\lambda_j,\theta),$$

- for $m = \lfloor (n-1)/2 \rfloor$, Whittle's M-estimator $\hat{\theta}_n$ is the solution to

$$\sum_{j=1}^{m} \psi_j(I(\lambda_j),\hat{\theta}_n) = 0.$$

- define $\mathcal{K}^*_{\mathcal{V}_n}(\upsilon,s) = \sum_j K^*_{\psi_j}(\upsilon,s)$, where

$$K^*_{\psi_j}(\upsilon,s) = \ln\left(E^*\left[\exp\{\upsilon\psi_j(I(\lambda_j),s)\}\right]\right),$$

with E^* computed treating $I(\lambda_j)/f(\lambda_j,\theta^0) \sim \exp^*(1)$.

Saddlepoint approximation (exponential-based)

The saddlepoint density approximation is:

$$g_{\hat{\theta}_n}(s) = \left[\frac{n}{2\pi \, \mathcal{K}^{*\,\prime\prime}_{\mathcal{V}_n}(\upsilon_0, s)} \right]^{1/2} e^{\mathcal{K}^{*}_{\mathcal{V}_n}(\upsilon_0, s)}, \tag{7}$$

and the saddlepoint $\upsilon_0 = \upsilon_0(s)$ solves

$$\mathcal{K}^{*\,\prime}_{\mathcal{V}_n}(\upsilon, s) = 0.$$

Remark

The advantage of using $I(\lambda)/f(\lambda, \theta) \sim \exp^*(1)$ is that $\mathcal{K}^{*}_{\mathcal{V}_n}$ is strictly convex, thus the saddlepoint equation admits a unique solution—which can be computed using standard methods, like the one based on the secant.

Saddlepoint approximation (exponential-based)

Example (ASY and frequency domain bootstrap (FDB))

Let us consider the AR(1):

$$Y_t = \theta_0 Y_{t-1} + \epsilon_t, \quad \epsilon_t \sim t_6 \quad \theta^0 = 0.4.$$

and the Whittle's estimator $\hat{\theta}_n$. Goal: approximate $P_{\theta^0}(\hat{\theta}_n > t_0)$ (e.g. for p-value).

Saddlepoint approximation (exponential-based)

Example (ASY and frequency domain bootstrap (FDB))

Let us consider the AR(1):

$$Y_t = \theta_0 Y_{t-1} + \epsilon_t, \quad \epsilon_t \sim t_6 \quad \theta^0 = 0.4.$$

and the Whittle's estimator $\hat{\theta}_n$. Goal: approximate $P_{\theta^0}(\hat{\theta}_n > t_0)$ (e.g. for p-value).

	12.5%	10%	5%	2.5%

Saddlepoint approximation (exponential-based)

Example (ASY and frequency domain bootstrap (FDB))

Let us consider the AR(1):

$$Y_t = \theta_0 Y_{t-1} + \epsilon_t, \quad \epsilon_t \sim t_6 \quad \theta^0 = 0.4.$$

and the Whittle's estimator $\hat{\theta}_n$. Goal: approximate $P_{\theta^0}(\hat{\theta}_n > t_0)$ (e.g. for p-value).

	12.5%	10%	5%	2.5%
		$n = 36$		
SAD	12.2%	9.1%	4.4%	2.0%
ASY	15.0%	11.8%	6.4%	3.2%
FDB	—	—	—	—

Saddlepoint approximation (exponential-based)

Example (ASY and frequency domain bootstrap (FDB))

Let us consider the AR(1):

$$Y_t = \theta_0 Y_{t-1} + \epsilon_t, \quad \epsilon_t \sim t_6 \quad \theta^0 = 0.4.$$

and the Whittle's estimator $\hat{\theta}_n$. Goal: approximate $P_{\theta^0}(\hat{\theta}_n > t_0)$ (e.g. for p-value).

	12.5%	10%	5%	2.5%
		$n = 36$		
SAD	12.2%	9.1%	4.4%	2.0%
ASY	15.0%	11.8%	6.4%	3.2%
FDB	—	—	—	—
		$n = 150$		
SAD	12.7%	9.9%	4.9%	2.3%
ASY	12.1%	9.2%	4.4%	2.0%
FDB	13.5%	10.8%	5.6%	2.9%
$(q_1; q_3)$	(10.5%; 15.7%)	(8.0%; 12.7%)	(4.0%; 6.6%)	(2.0%; 3.5%)

Saddlepoint approximation (exponential-based)

More generally, let $\theta = (\theta^{(1)}, \theta^{(2)})$, where $\theta^{(2)} \in \mathbb{R}^{p_2}, 1 < p_2 < p$ and consider testing

$$\mathcal{H}_0 : \theta^{(2)} = 0 \text{ vs } \mathcal{H}_1 : \theta^{(2)} > 0$$

with $\theta^{(1)}$ being the nuisance parameter (care is needed for σ^2).

Saddlepoint approximation (exponential-based)

More generally, let $\theta = (\theta^{(1)}, \theta^{(2)})$, where $\theta^{(2)} \in \mathbb{R}^{p_2}, 1 < p_2 < p$ and consider testing

$$\mathcal{H}_0 : \theta^{(2)} = 0 \;\; \text{vs} \;\; \mathcal{H}_1 : \theta^{(2)} > 0$$

with $\theta^{(1)}$ being the nuisance parameter (care is needed for σ^2). Two options:

Saddlepoint approximation (exponential-based)

More generally, let $\theta = (\theta^{(1)}, \theta^{(2)})$, where $\theta^{(2)} \in \mathbb{R}^{p_2}, 1 < p_2 < p$ and consider testing

$$\mathcal{H}_0 : \theta^{(2)} = 0 \;\text{ vs }\; \mathcal{H}_1 : \theta^{(2)} > 0$$

with $\theta^{(1)}$ being the nuisance parameter (care is needed for σ^2). Two options:

- $g_{\hat{\theta}_n}$ is available: construct the test using analytical marginalization techniques

Saddlepoint approximation (exponential-based)

More generally, let $\theta = (\theta^{(1)}, \theta^{(2)})$, where $\theta^{(2)} \in \mathbb{R}^{p_2}, 1 < p_2 < p$ and consider testing

$$\mathcal{H}_0 : \theta^{(2)} = 0 \quad \text{vs} \quad \mathcal{H}_1 : \theta^{(2)} > 0$$

with $\theta^{(1)}$ being the nuisance parameter (care is needed for σ^2). Two options:

- $g_{\hat{\theta}_n}$ is available: construct the test using analytical marginalization techniques
- adapt the univariate saddlepoint test statistic of Robinson et al (2003. AoS) which makes use of the approx c.g.f.:

$$\tilde{\mathcal{K}}^{\dagger}(\hat{\theta}_n^{(2)}) = 2 \inf_{\theta^{(1)}} \left[\sup_{\upsilon} \{ - \sum_j K^*_{\psi_j}(\upsilon; (\theta^{(1)}, \hat{\theta}_n^{(2)})) \} \right],$$

where υ solves the saddlepoint equation. The distribution of $\tilde{\mathcal{K}}^{\dagger}(\hat{\theta}_n^{(2)})$ under the null, can be approximated by a $\chi^2_{p_2}$ and the test statistic is asymptotically first order equivalent to the Wald test statistic.

Saddlepoint approximation (exponential-based)

Example (Gaussian ARFIMA $(0, d, 0)$)

Testing about the long-memory (simple hypothesis) for $n = 100, 250$:

$$\mathcal{H}_0 : d = d^0 \quad \text{vs} \quad \mathcal{H}_1 : d > d^0.$$

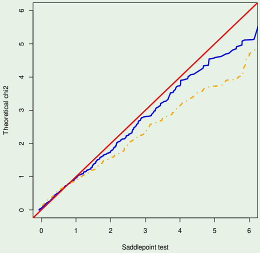
Saddlepoint approximation (exponential-based)

Example (Gaussian ARFIMA $(0, d, 0)$)

Testing about the long-memory (simple hypothesis) for $n = 100, 250$:

$$\mathcal{H}_0 : d = d^0 \quad \text{vs} \quad \mathcal{H}_1 : d > d^0.$$

$d^0 = 0.1$ $\qquad$ $d^0 = 0.35$ $\qquad$ Power $\left(n = 100, 250, 500\right)$



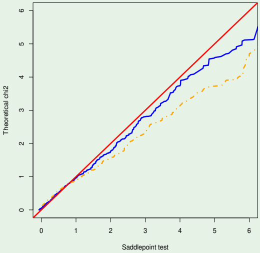
Saddlepoint approximation (exponential-based)

Example (Gaussian ARFIMA $(0, d, 0)$)

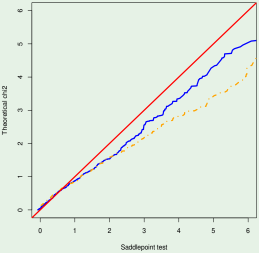
Testing about the long-memory (simple hypothesis) for $n = 100, 250$:

$$\mathcal{H}_0 : d = d^0 \quad \text{vs} \quad \mathcal{H}_1 : d > d^0.$$

$d^0 = 0.1$



$d^0 = 0.35$



Power ($n = 100, 250, 500$)

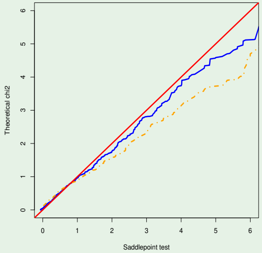
Saddlepoint approximation (exponential-based)

Example (Gaussian ARFIMA $(0, d, 0)$)

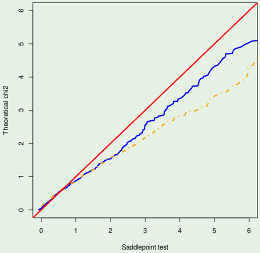
Testing about the long-memory (simple hypothesis) for $n = 100, 250$:

$$\mathcal{H}_0 : d = d^0 \quad \text{vs} \quad \mathcal{H}_1 : d > d^0.$$

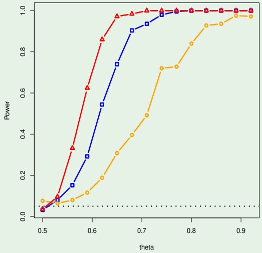
$d^0 = 0.1$



$d^0 = 0.35$



Power $(n = 100, 250, 500)$



Saddlepoint approximation (empirical version)

Remark

In the time series literature, some papers made use of empirical distribution of the periodogram ordinates: one may keep their independence but does not rely on the exponential distribution.

Saddlepoint approximation (empirical version)

Remark

In the time series literature, some papers made use of empirical distribution of the periodogram ordinates: one may keep their independence but does not rely on the exponential distribution.

- *Dahlhaus & Janas (1996. AoS) (FDB)*
- *Monti (1997, Biom.) (FDEL) and Nordman & Lahiri (2006, AoS)*
- *Kakizawa (2013, JTSA) (FDGEL)*

Saddlepoint approximation (empirical version)

Remark

In the time series literature, some papers made use of empirical distribution of the periodogram ordinates: one may keep their independence but does not rely on the exponential distribution.

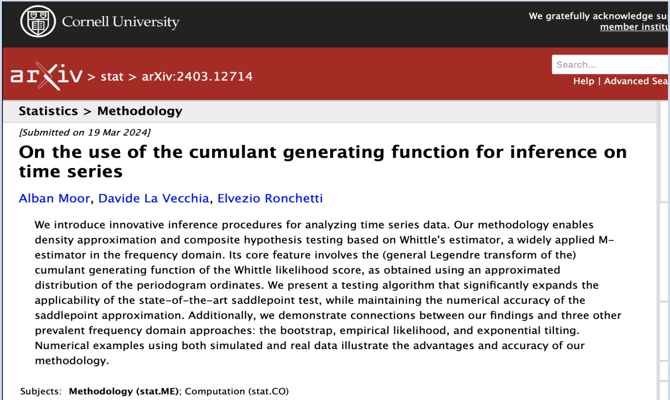
- *Dahlhaus & Janas (1996. AoS) (FDB)*
- *Monti (1997, Biom.) (FDEL) and Nordman & Lahiri (2006, AoS)*
- *Kakizawa (2013, JTSA) (FDGEL)*

This suggests that the c.g.f. may be approximated using the empirical distribution of the periodogram ordinates and obtain novel saddlepoint techniques!

Saddlepoint approximation (empirical version)

Remark

In the time series literature, some papers made use of empirical distribution of the periodogram ordinates: one may keep their independence but does not rely on the exponential distribution.



Cornell University

We gratefully acknowledge su
member institu

arXiv > stat > arXiv:2403.12714

Search...

Help | Advanced Sea

Statistics > Methodology

[Submitted on 19 Mar 2024]

On the use of the cumulant generating function for inference on time series

Alban Moor, Davide La Vecchia, Elvezio Ronchetti

We introduce innovative inference procedures for analyzing time series data. Our methodology enables density approximation and composite hypothesis testing based on Whittle's estimator, a widely applied M-estimator in the frequency domain. Its core feature involves the (general Legendre transform of the) cumulant generating function of the Whittle likelihood score, as obtained using an approximated distribution of the periodogram ordinates. We present a testing algorithm that significantly expands the applicability of the state-of-the-art saddlepoint test, while maintaining the numerical accuracy of the saddlepoint approximation. Additionally, we demonstrate connections between our findings and three other prevalent frequency domain approaches: the bootstrap, empirical likelihood, and exponential tilting. Numerical examples using both simulated and real data illustrate the advantages and accuracy of our methodology.

Subjects: **Methodology (stat.ME)**; Computation (stat.CO)

Saddlepoint approximation (empirical version)

The empirical saddlepoint density approximation is

$$\hat{g}_{\hat{\theta}_n}(s) = \left(\frac{m}{2\pi}\right)^{p/2} \left|\det \hat{M}(s)\right| \left|\det \hat{\Sigma}(s)\right|^{-1/2} \exp\{m \boxed{\hat{K}(s)}\}, \tag{8}$$

where

$$\boxed{\hat{K}(s)} = \boxed{\hat{K}(\hat{\upsilon}, s)} = \ln\left[\frac{1}{m}\sum_{j=1}^{m} \exp\{\hat{\upsilon}^T \psi_j(I_j, s)\}\right], \tag{9}$$

$$\hat{M}(s) = \frac{1}{m}\exp\{-\hat{K}(s)\}\sum_{j=1}^{m} \nabla_w \psi_j(I_j, w)|_{w=s} \exp\{\hat{\upsilon}^T \psi_j(I_j, s)\},$$

$$\hat{\Sigma}(s) = \frac{1}{m}\exp\{-\hat{K}(s)\}\sum_{j=1}^{m} \psi_j(I_j, s)\psi_j(I_j, s)^T \exp\{\hat{\upsilon}^T \psi_j(I_j, s)\}$$

and the empirical saddlepoint $\hat{\upsilon}$ satisfies:

$$\sum_{j=1}^{m} \psi_j(I_j, s) \exp\{\hat{\upsilon}^T \psi_j(I_j, s)\} = 0. \tag{10}$$

Saddlepoint approximation (empirical version)

The empirical saddlepoint is based on the c.g.f. $\boxed{\hat{K}}$ as an approximation to the true c.g.f.: it is the key tool needed to compute $\hat{g}_{\hat{\theta}_n}$ and it unveils important connection with the FDEL (Monti (1997, Biom.)).

Saddlepoint approximation (empirical version)

The empirical saddlepoint is based on the c.g.f. $\hat{K}$ as an approximation to the true c.g.f.: it is the key tool needed to compute $\hat{g}_{\hat{\theta}_n}$ and it unveils important connection with the FDEL (Monti (1997, Biom.)).
Indeed, FDEL solves the system of (tilted) estimating equations

$$\sum_{j=1}^{m} \psi_j(I_j, s)[1 + \hat{\xi}^T \psi_j(I_j; s)]^{-1} = 0, \tag{11}$$

where we use the shorthand notation $\hat{\xi} = \hat{\xi}(s)$. Then, Monti defines a FD version of Owen's statistics as

$$\hat{W}(s) = 2 \sum_{j=1}^{m} \ln\{1 + \hat{\xi}^T \psi_j(I_j; s)\},$$

and $\hat{W}$ is first-order equivalent to W_n.

Saddlepoint approximation (empirical version)

Now notice that

- the saddlepoint satisfies (Taylor expansion of the exp) the equation

$$\sum_{j=1}^{m} \psi_j(I_j; s)[1 + \hat{\upsilon}^T \psi_j(I_j; s)] = O_P(n^{-1}),$$

since $\hat{\upsilon} = O_P(n^{-1/2})$.

Saddlepoint approximation (empirical version)

Now notice that

- the saddlepoint satisfies (Taylor expansion of the exp) the equation

$$\sum_{j=1}^{m} \psi_j(I_j; s)[1 + \hat{\upsilon}^T \psi_j(I_j; s)] = O_P(n^{-1}),$$

  since $\hat{\upsilon} = O_P(n^{-1/2})$.
- a Taylor expansion of the equation defining the FDEL yields

$$\sum_{j=1}^{m} \psi_j(I_j; s)[1 - \hat{\xi}^T \psi_j(I_j; s)] = O_P(n^{-1}),$$

  since $\hat{\xi} = O_P(n^{-1/2})$.

Saddlepoint approximation (empirical version)

Now notice that

- the saddlepoint satisfies (Taylor expansion of the exp) the equation

$$\sum_{j=1}^{m} \psi_j(I_j; s)[1 + \hat{\upsilon}^T \psi_j(I_j; s)] = O_P(n^{-1}),$$

  since $\hat{\upsilon} = O_P(n^{-1/2})$.
- a Taylor expansion of the equation defining the FDEL yields

$$\sum_{j=1}^{m} \psi_j(I_j; s)[1 - \hat{\xi}^T \psi_j(I_j; s)] = O_P(n^{-1}),$$

  since $\hat{\xi} = O_P(n^{-1/2})$.

Remark

The empirical saddlepoint and the empirical likelihood solve at the order $O_P(n^{-1})$ the same equation.

Saddlepoint approximation (empirical version)

Building on this remark, we prove that:

$$-2n \underbrace{\boxed{\hat{K}(s)}}_{\text{Emp Sadd Test}} = 2 \underbrace{\hat{W}(s)}_{\text{Owen stat}} - \frac{2m^{-1/2}}{3} \sum_{j=1}^{m} \left\{ u^T \hat{M}^T \hat{\Sigma}^{-1} \psi_j(I_j; \hat{\theta}_n) \right\}^3 + R_n$$

where, under some conditions, $R_n = O_P(n^{-1})$, $\hat{\Sigma} = \hat{\Sigma}(\hat{\theta}_n)$ and $\hat{M} = \hat{M}(\hat{\theta}_n)$.

Saddlepoint approximation (empirical version)

Building on this remark, we prove that:

$$-2n \underbrace{\hat{K}(s)}_{\text{Emp Sadd Test}} = 2 \underbrace{\hat{W}(s)}_{\text{Owen stat}} - \frac{2m^{-1/2}}{3} \sum_{j=1}^{m} \left\{ u^T \hat{M}^T \hat{\Sigma}^{-1} \psi_j(I_j; \hat{\theta}_n) \right\}^3 + R_n$$

where, under some conditions, $R_n = O_P(n^{-1})$, $\hat{\Sigma} = \hat{\Sigma}(\hat{\theta}_n)$ and $\hat{M} = \hat{M}(\hat{\theta}_n)$.

Remark

The latter result has a three-fold theoretical importance:

(i) *it connects our FDES to the FDEL*

Saddlepoint approximation (empirical version)

Building on this remark, we prove that:

$$-2n \underbrace{\boxed{\hat{K}(s)}}_{\text{Emp Sadd Test}} = 2 \underbrace{\hat{W}(s)}_{\text{Owen stat}} - \frac{2m^{-1/2}}{3} \sum_{j=1}^{m} \left\{ u^T \hat{M}^T \hat{\Sigma}^{-1} \psi_j(I_j; \hat{\theta}_n) \right\}^3 + R_n$$

where, under some conditions, $R_n = O_P(n^{-1})$, $\hat{\Sigma} = \hat{\Sigma}(\hat{\theta}_n)$ and $\hat{M} = \hat{M}(\hat{\theta}_n)$.

Remark

The latter result has a three-fold theoretical importance:

(i) *it connects our FDES to the FDEL*

(ii) *it illustrates that the difference between* $\boxed{\hat{K}}$ *and* $\hat{W}$ *depends on the third moment of the Whittle's score: both correct* W_n *for the skewness but in a different way*

Saddlepoint approximation (empirical version)

Building on this remark, we prove that:

$$-2n \underbrace{\boxed{\hat{K}(s)}}_{\text{Emp Sadd Test}} = 2 \underbrace{\hat{W}(s)}_{\text{Owen stat}} - \frac{2m^{-1/2}}{3} \sum_{j=1}^{m} \left\{ u^T \hat{M}^T \hat{\Sigma}^{-1} \psi_j(I_j; \hat{\theta}_n) \right\}^3 + R_n$$

where, under some conditions, $R_n = O_P(n^{-1})$, $\hat{\Sigma} = \hat{\Sigma}(\hat{\theta}_n)$ and $\hat{M} = \hat{M}(\hat{\theta}_n)$.

Remark

The latter result has a three-fold theoretical importance:

(i) *it connects our FDES to the FDEL*

(ii) *it illustrates that the difference between $\boxed{\hat{K}}$ and $\hat{W}$ depends on the third moment of the Whittle's score: both correct W_n for the skewness but in a different way*

(iii) *it yields a nonparametric approximation to the density of Whittle's estimator based on the FDEL*

Saddlepoint approximation (empirical version)

On the practical side: use the empirical saddlepoint under $\mathcal{H}_0$ to approximate the distribution of W_n (or EL, ET) test statistics, where

$$\mathcal{H}_0 : \theta = \theta^0 \text{ vs. } \mathcal{H}_1 : \theta \neq \theta^0.$$

Saddlepoint approximation (empirical version)

On the practical side: use the empirical saddlepoint under $\mathcal{H}_0$ to approximate the distribution of W_n (or EL, ET) test statistics, where

$$\mathcal{H}_0 : \theta = \theta^0 \text{ vs. } \mathcal{H}_1 : \theta \neq \theta^0.$$

To this end,

- We define the Wald-type statistic, with $\hat{V} = \hat{M}^{-1}\hat{\Sigma}\hat{M}^{-1}$ (estimate of asym var of Whittle estim.),

$$W_n(\theta) = n(\hat{\theta}_n - \theta)^T \hat{V}^{-1}(\hat{\theta}_n - \theta) \overset{\text{asy}}{\sim} \chi^2_p, \;\; p = \dim(\theta)$$

- In contrast, we make use of $\hat{g}_{\hat{\theta}_n}$ to obtain the approximated p-value

$$P[W_n(\theta^0) > \tilde{w}(\theta^0) \mid \mathcal{H}_0] \approx 1 - \int_{\mathcal{B}} \hat{g}_{\hat{\theta}_n}(\theta) d\theta, \tag{12}$$

where $\mathcal{B} = \left\{\theta \in \mathbb{R}^p \mid \tilde{w}(\theta) \geq \tilde{w}(\theta^0)\right\}$, with $\tilde{w}(\theta^0)$ being the observed value of the test statistic at θ^0.

Saddlepoint approximation (empirical version)

On the practical side: use the empirical saddlepoint under $\mathcal{H}_0$ to approximate the distribution of W_n (or EL, ET) test statistics, where

$$\mathcal{H}_0 : \theta = \theta^0 \text{ vs. } \mathcal{H}_1 : \theta \neq \theta^0.$$

To this end,

- We define the Wald-type statistic, with $\hat{V} = \hat{M}^{-1}\hat{\Sigma}\hat{M}^{-1}$ (estimate of asym var of Whittle estim.),

$$W_n(\theta) = n(\hat{\theta}_n - \theta)^T \hat{V}^{-1}(\hat{\theta}_n - \theta) \overset{\text{asy}}{\sim} \chi^2_p, \;\; p = \dim(\theta)$$

- In contrast, we make use of $\hat{g}_{\hat{\theta}_n}$ to obtain the approximated p-value

$$P[W_n(\theta^0) > \tilde{w}(\theta^0) \mid \mathcal{H}_0] \approx 1 - \int_{\mathcal{B}} \hat{g}_{\hat{\theta}_n}(\theta) d\theta, \qquad (12)$$

  where $\mathcal{B} = \left\{\theta \in \mathbb{R}^p \mid \tilde{w}(\theta) \geq \tilde{w}(\theta^0)\right\}$, with $\tilde{w}(\theta^0)$ being the observed value of the test statistic at θ^0.

- **To compute the integral in (12), we suggest to use an importance sampling scheme** based on an instrumental Gaussian distribution.

Saddlepoint approximation (empirical version)

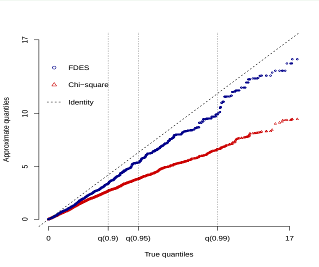
Example

We consider an **ARFIMA**$(0, d, 0)$ model test for

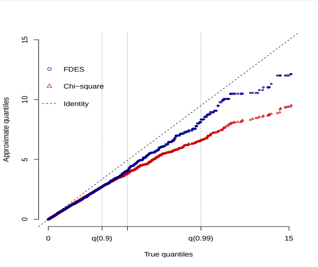
$$\mathcal{H}_0 : \; d^0 = 0 \text{ vs. } \mathcal{H}_1 : \; d \neq 0$$

using the empirical saddlepoint. We compare the approx quantiles to true quantiles (as obtained by MC simulations), for the saddlepoint technique and first-order asymptotic theory (χ^2_1).

$n = 30$



$n = 250$



Saddlepoint approximation (empirical version)

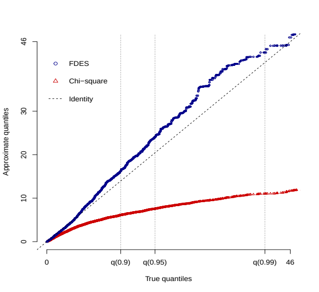
Example

We consider an **ARFIMA(1,d,1)** with $\theta^0 = (0.5, 0.25, 0.5)$ and test

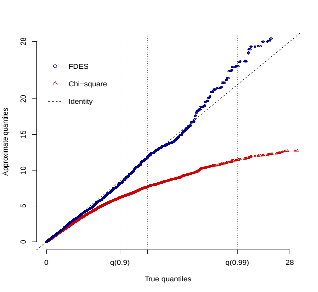
$$\mathcal{H}_0 : \theta = \theta^0 \text{ vs. } \mathcal{H}_1 : \theta \neq \theta^0$$

using the empirical saddlepoint. We compare the approx quantiles to true quantiles (as obtained by MC simulations), for the saddlepoint technique and first-order asymptotic theory (χ^2_3).

$n = 100$



$n = 500$



Saddlepoint approximation (empirical version)

Remark

- **Testing in the presence of nuisance.** *There are cases where only certain components of* θ *have to be tested. Namely, taking the partition* $(\theta_{(1)}\ \theta_{(2)})$, *we test*

$$\mathcal{H}_0 : \theta_{(1)} = \theta^0_{(1)}$$

w.l.o.g. To perform this type of test, we can simply modify our procedure, redefining the integration set as

$$\mathcal{B} \leftarrow \left\{\theta \in \Theta \mid \tilde{w}\left(\theta_{(1)}, \hat{\theta}_{(2),n}\right) > \tilde{w}\left(\theta^0_{(1)}, \hat{\theta}_{(2),n}\right)\right\},$$

instead of

$$\mathcal{B} \leftarrow \left\{\theta \in \Theta \mid \tilde{w}(\theta) > \tilde{w}(\theta^0)\right\}.$$

$\Rightarrow$ *our numerical integration via importance sampling avoids to take the infimum w.r.t. to the nuisance parameters.*

Saddlepoint approximation (empirical version)

Remark

- **Testing in the presence of nuisance.** *There are cases where only certain components of* θ *have to be tested. Namely, taking the partition* $(\theta_{(1)}\ \theta_{(2)})$, *we test*

$$\mathcal{H}_0 : \theta_{(1)} = \theta_{(1)}^0$$

w.l.o.g. To perform this type of test, we can simply modify our procedure, redefining the integration set as

$$\mathcal{B} \leftarrow \left\{\theta \in \Theta \mid \tilde{w}\left(\theta_{(1)}, \hat{\theta}_{(2),n}\right) > \tilde{w}\left(\theta_{(1)}^0, \hat{\theta}_{(2),n}\right)\right\},$$

instead of

$$\mathcal{B} \leftarrow \left\{\theta \in \Theta \mid \tilde{w}(\theta) > \tilde{w}(\theta^0)\right\}.$$

$\Rightarrow$ *our numerical integration via importance sampling avoids to take the infimum w.r.t. to the nuisance parameters.*

- **Numerical accuracy.** *Also using the empirical distribution of the periodogram ordinates, the saddlepoint technique yields an improvement on the first order asymptotic theory.*

Thank you

For questions: davide.lavecchia@unige.ch

Laplace in brief

The Laplace method is typically applied to approximate integrals of type:

$$\int_a^b e^{v \, k(x)} dx,$$

where $k(\cdot)$ has unique maximum at $x_0 \in (a, b) \subset \mathbb{R}$ and $v \in \mathbb{R}^+$ is large.

Laplace in brief

The Laplace method is typically applied to approximate integrals of type:

$$\int_a^b e^{v \boxed{k}(x)} dx,$$

where $k(\cdot)$ has unique maximum at $x_0 \in (a, b) \subset \mathbb{R}$ and $v \in \mathbb{R}^+$ is large. A second-order Taylor expansion for $k(\cdot)$ yields

$$\int_a^b e^{v \boxed{k}(x)} dx \sim e^{v \boxed{k}(x_0)} \int_{x_0-\epsilon}^{x_0+\epsilon} e^{v \boxed{k^{('')}}(x_0) \frac{x^2}{2}} dx$$

Laplace in brief

The Laplace method is typically applied to approximate integrals of type:

$$\int_a^b e^{v \boxed{k}(x)} dx,$$

where $k(\cdot)$ has unique maximum at $x_0 \in (a, b) \subset \mathbb{R}$ and $v \in \mathbb{R}^+$ is large. A second-order Taylor expansion for $k(\cdot)$ yields

$$\int_a^b e^{v \boxed{k}(x)} dx \sim e^{v \boxed{k}(x_0)} \int_{x_0-\epsilon}^{x_0+\epsilon} e^{v \boxed{k^{('')}}(x_0)\frac{x^2}{2}} dx \sim e^{v \boxed{k}(x_0)} \sqrt{\frac{2\pi}{-v \boxed{k^{('')}}(x_0)}},$$

Laplace in brief

The Laplace method is typically applied to approximate integrals of type:

$$\int_a^b e^{v\, k(x)} dx,$$

where $k(\cdot)$ has unique maximum at $x_0 \in (a, b) \subset \mathbb{R}$ and $v \in \mathbb{R}^+$ is large. A second-order Taylor expansion for $k(\cdot)$ yields

$$\int_a^b e^{v\, k(x)} dx \sim e^{v\, k(x_0)} \int_{x_0-\epsilon}^{x_0+\epsilon} e^{v\, k^{('')}(x_0)\frac{x^2}{2}} dx \sim e^{v\, k(x_0)} \sqrt{\frac{2\pi}{-v\, k^{('')}(x_0)}},$$

where (i) for $\epsilon > 0$, we deform the path of integration $\int_a^b \mapsto \int_{x_0-\epsilon}^{x_0+\epsilon}$

Laplace in brief

The Laplace method is typically applied to approximate integrals of type:

$$\int_a^b e^{v\, k(x)}\, dx,$$

where $k(\cdot)$ has unique maximum at $x_0 \in (a, b) \subset \mathbb{R}$ and $v \in \mathbb{R}^+$ is large. A second-order Taylor expansion for $k(\cdot)$ yields

$$\int_a^b e^{v\, k(x)} dx \sim e^{v\, k(x_0)} \int_{x_0-\epsilon}^{x_0+\epsilon} e^{v\, k^{('')}(x_0)\frac{x^2}{2}} dx \sim e^{v\, k(x_0)} \sqrt{\frac{2\pi}{-v\, k^{('')}(x_0)}},$$

where (i) for $\epsilon > 0$, we deform the path of integration $\int_a^b \mapsto \int_{x_0-\epsilon}^{x_0+\epsilon}$ and (ii) we solve the Gaussian integral—getting an approx featuring relative error, under suitable assumptions.

Jump Back