



# 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...



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

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



Source: Maxine's drawing book, copyright 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)

## First part: i.i.d. setting

## Motivation from theoretical statistics

Theorem (Karlin-Rubin, as stated in Casella-Berger)

Consider testing

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

Suppose that  $T$  is a sufficient statistic for  $\theta$  and the family of pdfs or pmfs  $\{g(t | \theta) : \theta \in \Theta\}$  of  $T$  has a Monotone Likelihood Ratio. Then for any  $t_0$ , the test that rejects  $\mathcal{H}_0$  if and only if  $T > t_0$  is a UMP level  $\alpha$  test, where

$$\alpha = P_{\theta^0}(T > t_0).$$

## Motivation from theoretical statistics

Theorem (Karlin-Rubin, as stated in Casella-Berger)

Consider testing

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

Suppose that  $T$  is a sufficient statistic for  $\theta$  and the family of pdfs or pmfs  $\{g(t | \theta) : \theta \in \Theta\}$  of  $T$  has a Monotone Likelihood Ratio. Then for any  $t_0$ , the test that rejects  $\mathcal{H}_0$  if and only if  $T > t_0$  is a UMP level  $\alpha$  test, where

$$\alpha = P_{\theta^0}(T > t_0).$$

We need to compute that probability, which is related to the CDF or pdf of a functional of the data!

# 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.

- ➊ 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.

- ➊ 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:

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

- ➋ 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.

- ➊ 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:

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

- ➋ 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)

- ➌ 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.

- ➊ 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:

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

- ➋ 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)

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

- ➊ 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:

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

- ➋ 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)

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

## Motivation from theoretical statistics

*Typical statistical problem:* For a given statistic/functional

$$T : \text{dom } T \rightarrow \mathbb{R}^p, p \geq 1$$

(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

$$T : \text{dom } T \rightarrow \mathbb{R}^p, p \geq 1$$

(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), **tail probabilities (more generally the whole distribution) cannot be computed exactly.**

## Motivation from theoretical statistics

*Typical statistical problem:* For a given statistic/functional

$$T : \text{dom } T \rightarrow \mathbb{R}^p, p \geq 1$$

(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), **tail probabilities (more generally the whole distribution) cannot be computed exactly.**

⇒ we have to rely on **approximations to  $P$**

## Motivation from theoretical statistics

We can approximate tail probabilities via

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

## Motivation from theoretical statistics

We can approximate 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

# Motivation from theoretical statistics

We can approximate 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

## Motivation from theoretical statistics

We can approximate 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

# 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	$\mu$	$\nu$
Charact. fct.	$\chi(u)$	$\xi(u)$
Cumulants	$\beta_r$	$\gamma_r$

where the two Fourier transforms are

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

and the cumulants are (by definition)

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

# Analytical techniques: Edgeworth expansion

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

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

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  $u = 0$** :

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

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  $F$  is  $G$ .

# Analytical techniques: Edgeworth expansion

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

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

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  $F$  is  $G$ .

4 Differentiation yields an expansion for the density of  $X$ .

# Motivation from 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  $\mathbf{X} = (X_1, \dots, X_n)$  of i.i.d. copies of  $X$ , whose mgf and cgf exist and:  $E_\mu[X] = 0$ ,  $V_\mu(X) = \sigma^2 < \infty$ .

# Motivation from 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  $\mathbf{X} = (X_1, \dots, X_n)$  of i.i.d. copies of  $X$ , whose mgf and cgf exist and:  $E_\mu[X] = 0$ ,  $V_\mu(X) = \sigma^2 < \infty$ .

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

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

# Motivation from 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  $\mathbf{X} = (X_1, \dots, X_n)$  of i.i.d. copies of  $X$ , whose mgf and cgf exist and:  $E_\mu[X] = 0$ ,  $V_\mu(X) = \sigma^2 < \infty$ .

The standardized mean (statistic,  $T(\mathbf{X})$ ) 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 from 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  $\mathbf{X} = (X_1, \dots, X_n)$  of i.i.d. copies of  $X$ , whose mgf and cgf exist.



BROWSE ▾

RESOURCES ▾

ABOUT ▾

Statist. Sci.

Vol.38 • No. 1 • February 2023

Home ▾

February 2023

### On Some Connections Between Esscher's Tilting Saddlepoint Approximations, and Optimal Transportation: A Statistical Perspective

Davide La Vecchia, Elvezio Ronchetti, Andrej Ilijevski

Author Affiliations +

Statist. Sci. 38(1): 30-51 (February 2023). DOI: 10.1214/21-STSB47



## Motivation from 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 normal density and higher order terms correct for skewness, kurtosis:

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

## Motivation from 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 normal density and higher order terms correct for skewness, kurtosis:

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

# Motivation from 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 normal density and higher order terms correct for skewness, kurtosis:

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

with  $\lambda_3$  and  $\lambda_4$  being the standardized cumulants of  $X$  of order three and four, while  $\phi$  is the pdf of a standard normal.

## Motivation from 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 normal density and higher order terms correct for skewness, kurtosis:

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

with  $\lambda_3$  and  $\lambda_4$  being the standardized cumulants of  $X$  of order three and four, while  $\phi$  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 from 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 normal density and higher order terms correct for skewness, kurtosis:

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

with  $\lambda_3$  and  $\lambda_4$  being the standardized cumulants of  $X$  of order three and four, while  $\phi$  is the pdf of a standard normal.

### Remark

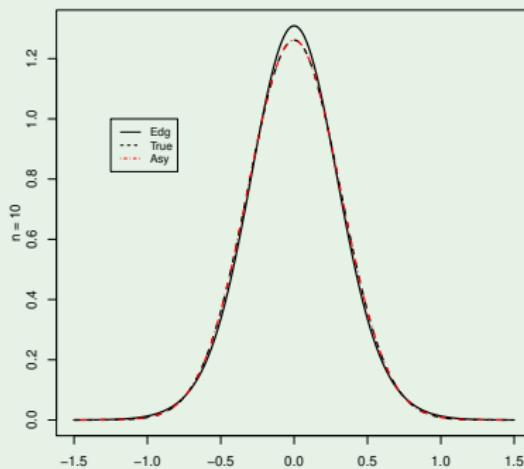
By construction, Edgeworth expansions provide in general a good approximation in the center of the density, BUT they can be inaccurate in the tails, where they can even become negative.

# Motivation from theoretical statistics

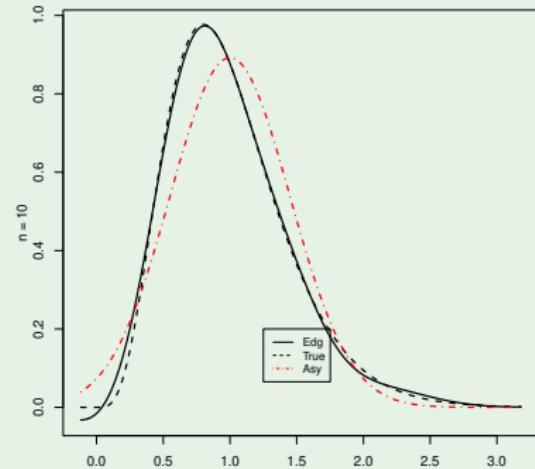
## Example (cont'd)

For Asy and Edg, consider  $\bar{X}_n$  for  $n = 10, 50, 250$ , for  $X_i \sim \mathcal{N}(0, 1)$  and  $X_i \sim \text{exp}(1)$

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



$\text{exp}(1)(n = 10)$

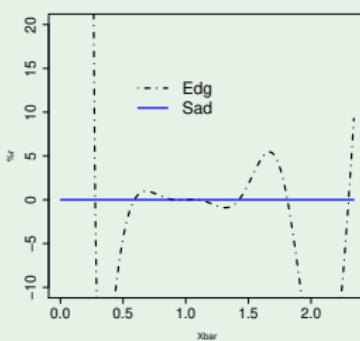


# Motivation from theoretical statistics

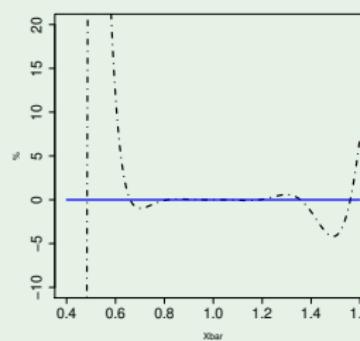
## Example (cont'd)

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

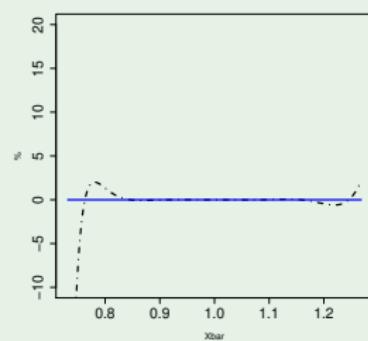
$n = 10$



$n = 50$



$n = 250$



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

# Motivation from 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)):

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

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

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

# Motivation from 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)):

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

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 from 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 from 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

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

# Motivation from 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  $\mathcal{N}(0, \frac{1}{n})$  it's exact!!

# Motivation from 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  $\mathcal{N}(0, \frac{1}{n})$  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 from 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  $\mathcal{N}(0, \frac{1}{n})$  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 from theoretical statistics

## Example (cont'd)

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

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

# Motivation from theoretical statistics

## Example (cont'd)

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

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

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

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

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

## Motivation from 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

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

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

## Motivation from theoretical statistics

The saddle 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

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

with  $\nu \in \mathbb{R}^+$  is large,  $\xi$  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  $w'(z)$ . Then, we approximate the resulting integral using a series expansion (Watson's lemma). See Daniels (AoMS, 1954).

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

[Jump to Laplace](#)

# Motivation from theoretical statistics

Alternative: derive  $g_n$  via convex analysis.

## Motivation from 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 from 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  $v(s)$ , 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  $v(s)$ )

# Motivation from 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  $v(s)$ , 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  $v(s)$ )
- compute a low-order **Edgeworth expansion** for the tilted density (centered at  $s$ , so it works well!) to obtain  $g_n(s)$

# Motivation from 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  $v(s)$ , 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  $v(s)$ )
- 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 from 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  $v(s)$ , 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  $v(s)$ )
- 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  $s \in \mathbb{R}$

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

## Motivations related to dependent data analysis

Many macroeconomic time series display a persistent time trend and contain only a few observations recorded at annual frequency. Much controversy in macroeconomics 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 time series display a persistent time trend and contain only a few observations recorded at annual frequency. Much controversy in macroeconomics 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 time series** display a persistent time trend and contain only a few observations recorded at annual frequency. Much controversy in macroeconomics 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



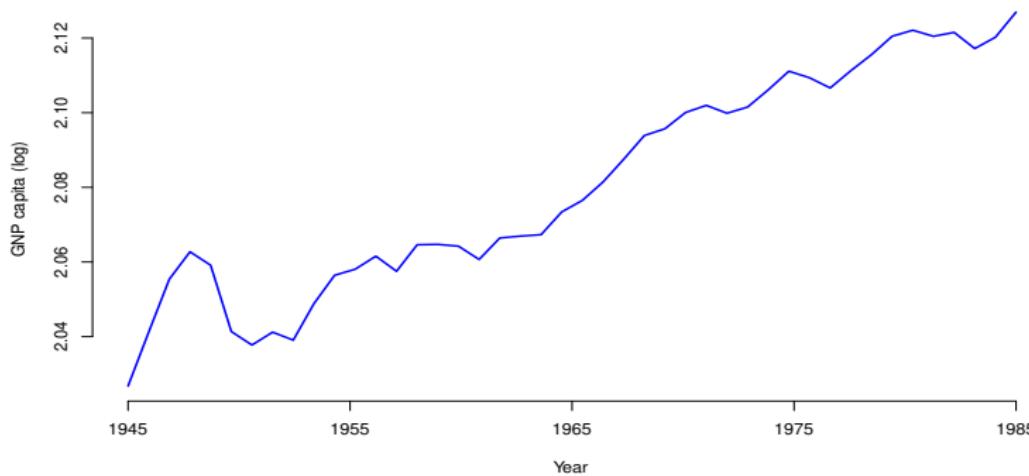
---

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

Davide La Vecchia   , Elvezio Ronchetti

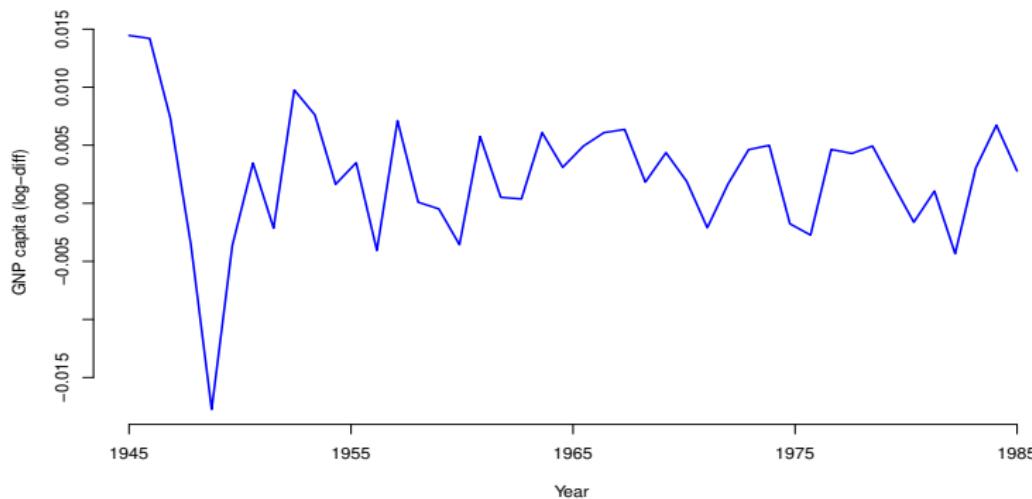
## Motivations related to dependent data analysis

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

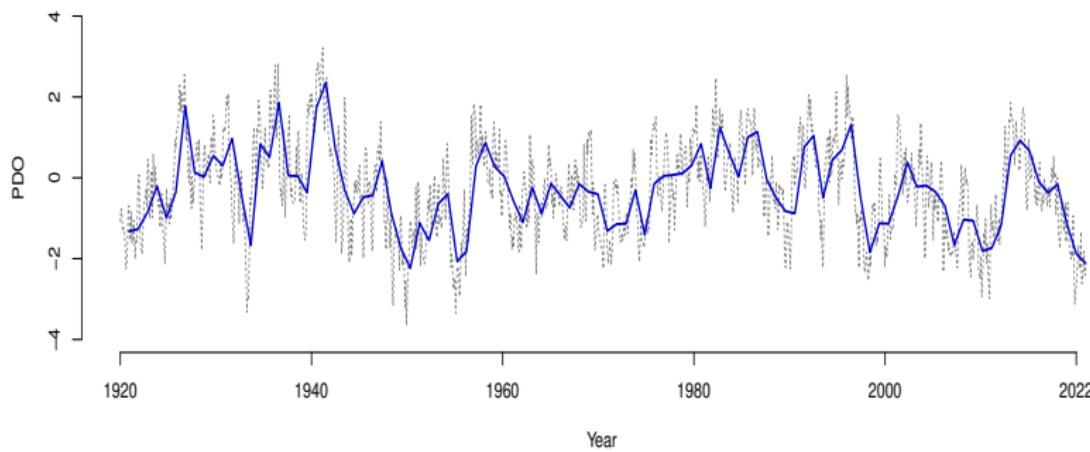
In the literature one is typically testing for the presence of long memory: ARFIMA models and

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

we resort on an M-estimator (Whittle), which is *asymptotically*  $\chi^2$  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

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

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

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

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

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

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

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

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  $\theta$  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

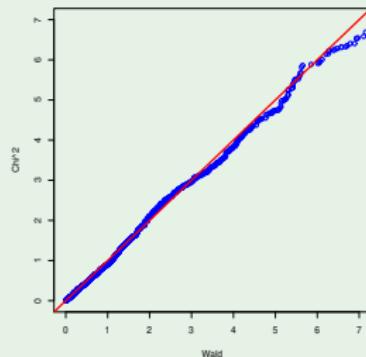
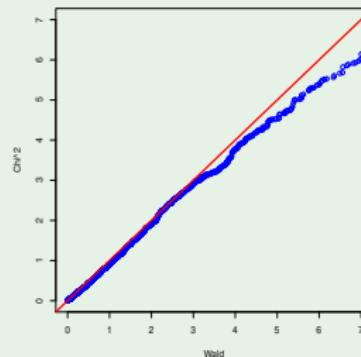
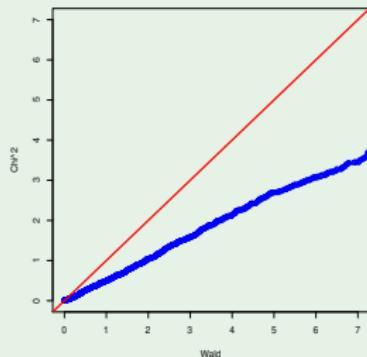
$$\mathcal{H}_0 : d = 0 \text{ vs. } \mathcal{H}_1 : d > 0,$$

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

$n = 250$

$n = 2500$

$n = 5000$



# 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  $\chi^2$  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:

The screenshot shows the homepage of the Journal of the American Statistical Association. The top navigation bar includes links for Home, All Journals, Journal of the American Statistical Association, List of Issues, Latest Articles, Saddlepoint Approximations for Spatial P ..., and a search bar. Below the navigation is the journal logo and a menu with options to Submit an article and Journal homepage. A large green arrow points down to the article details. The article title is "Saddlepoint Approximations for Spatial Panel Data Models" by Chaonan Jiang, Davide La Vecchia, Elvezio Ronchetti & Olivier Scaillet. It was received on 01 Jul 2020, accepted on 09 Sep 2021, and published online on 20 Sep 2021. The article has a DOI of <https://doi.org/10.1080/01621459.2021.1981913>. There are also links for "Download citation" and "Check for updates". On the right side of the article details, there is a "Full access" button with a checkmark icon.

## **Second part:** time series setting

# 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 & Kakizawa (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 & Kakizawa (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

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

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

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

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

Under suitable assumptions, we have (for  $i \in \mathbb{C}$ )

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

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

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

# Some elements of spectral analysis

## Definition

Given time series data  $Y_1, \dots, Y_n$ , 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  $j = 0, 1, \dots, n - 1$ , where the frequencies  $\lambda_j = j/n$  are called Fourier or fundamental frequencies.

# Some elements of spectral analysis

## Definition

Given time series data  $Y_1, \dots, Y_n$ , 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  $j = 0, 1, \dots, n - 1$ , where the frequencies  $\lambda_j = j/n$  are called Fourier or fundamental frequencies.

The periodogram at  $\lambda_j$  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 covariance.

## Some elements of spectral analysis

**Property 1.** The periodogram is an asymptotically unbiased (nonparametric) estimator of the spectral density  $f(\lambda)$ .

## Some elements of spectral analysis

**Property 1.** The periodogram is an asymptotically unbiased (nonparametric) estimator of the spectral density  $f(\lambda)$ .

**Property 2.** The periodogram ordinates are such that

$$I(\lambda) \xrightarrow{\mathcal{D}} i.d. \xi f(\lambda), \quad \xi \sim \exp(1) \quad (4)$$

### 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

Property 2 allows to derive a frequency domain likelihood and parameter estimation is obtained maximizing this likelihood.

## Some elements of spectral analysis

Property 2 allows to derive a frequency domain likelihood and parameter estimation is obtained maximizing this likelihood.

This idea goes back to [Whittle \(1951\)](#): if there is a parametric model for  $f(\lambda, \theta)$ , then we may work on:

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

which is obtained making use of Property 2 ( $\lambda$  is in radians, from now on).

## Some elements of spectral analysis

Property 2 allows to derive a frequency domain likelihood and parameter estimation is obtained maximizing this likelihood.

This idea goes back to [Whittle \(1951\)](#): if there is a parametric model for  $f(\lambda, \theta)$ , then we may work on:

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

which is obtained making use of Property 2 ( $\lambda$  is in radians, from now on).

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

$$\hat{\theta}_n = \arg \max_{\theta} \mathcal{L}_W(\theta)$$

(or  $\nabla_{\theta} \mathcal{L}_W(\hat{\theta}_n) = 0$ ) defines an **M-estimator in the frequency domain**. Then,

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

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

## Setting: SRD and LRD

Suppose that  $\{Y_t\}$  is a **linear process**, second order stationary process,

$$Y_t = \sum_{r=0}^{\infty} a_r \varepsilon_{t-r},$$

for  $t \in \mathbb{Z}$ ,

## Setting: SRD and LRD

Suppose that  $\{Y_t\}$  is a **linear process**, second order stationary process,

$$Y_t = \sum_{r=0}^{\infty} a_r \varepsilon_{t-r},$$

for  $t \in \mathbb{Z}$ , with spectral density function

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

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

## Setting: SRD and LRD

Suppose that  $\{Y_t\}$  is a **linear process**, second order stationary process,

$$Y_t = \sum_{r=0}^{\infty} a_r \varepsilon_{t-r},$$

for  $t \in \mathbb{Z}$ , with spectral density function

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

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 implies

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

## Saddlepoint approximation (exponential-based)

First order asymptotic theory implies

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

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)

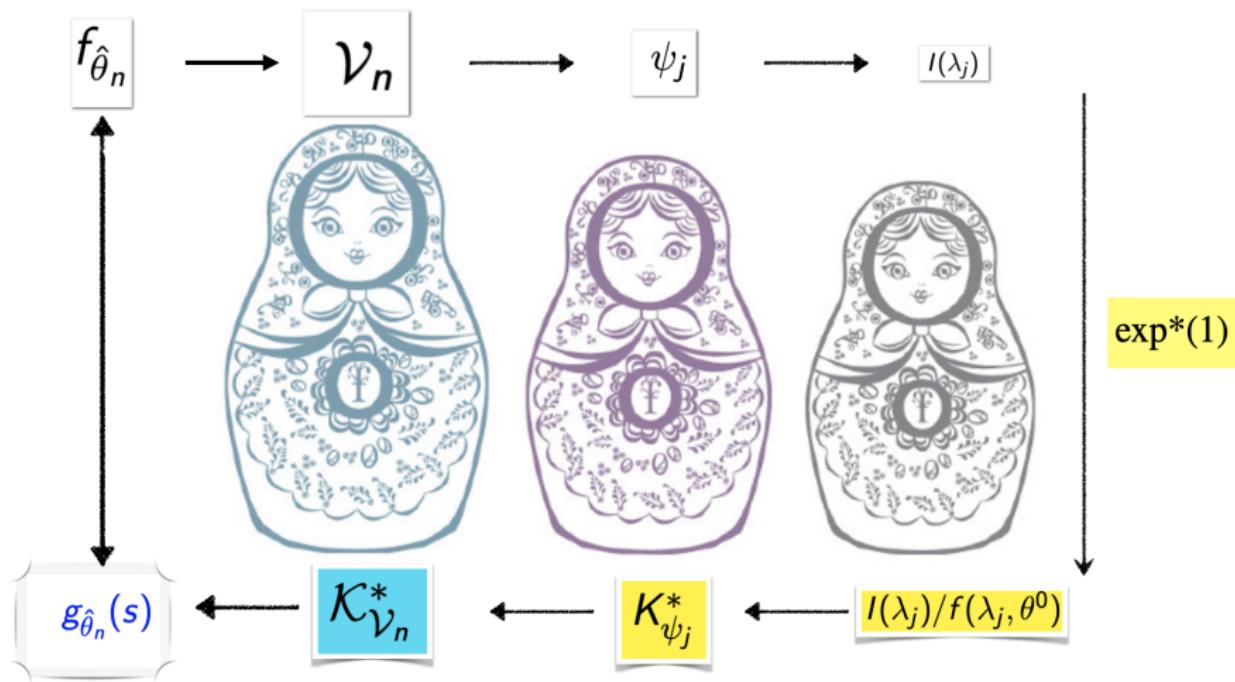
First order asymptotic theory implies

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

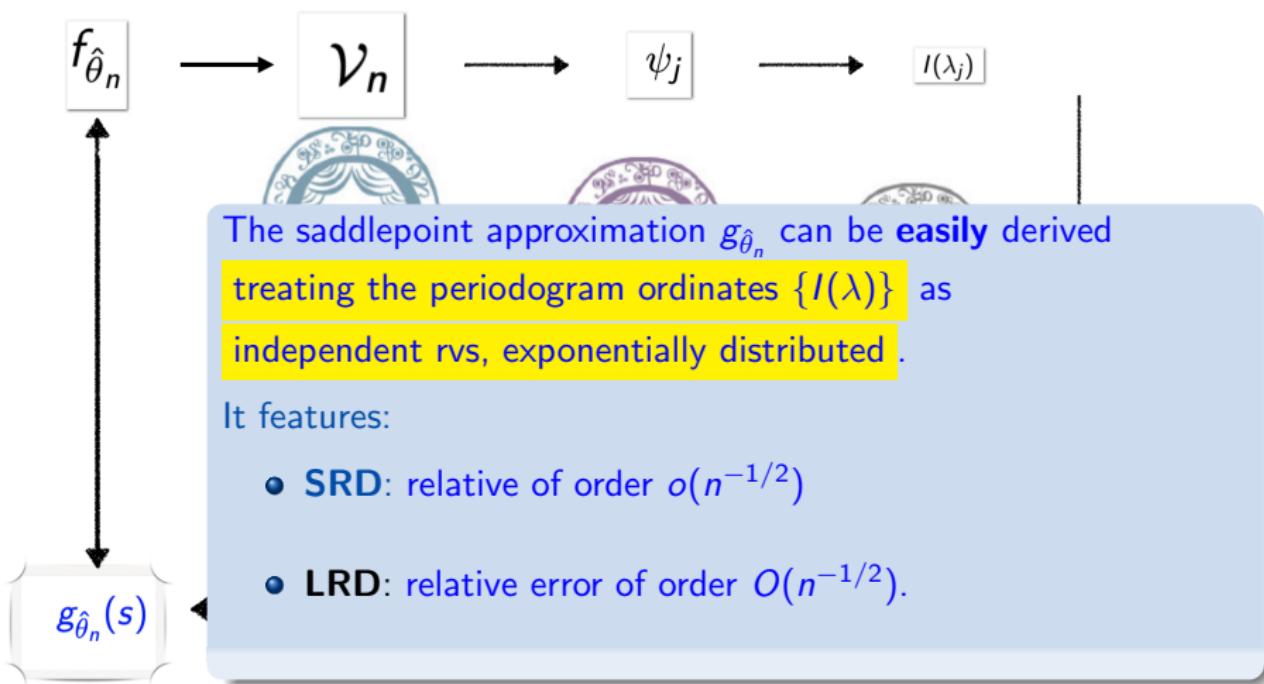
To have a better density approximation to  $f_{\hat{\theta}_n}$ , we may derive the saddlepoint density approximation  $g_{\hat{\theta}_n}$  treating the periodogram ordinates as independently and exponentially distributed r.v.'s: we use it to approximate the c.g.f.  $\mathcal{K}_{\mathcal{V}_n}$  ...

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

# Visualisation



# Visualisation



# Saddlepoint approximation (exponential-based)

Specifically:

- Whittle's estimating function is

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

## Saddlepoint approximation (exponential-based)

Specifically:

- Whittle's estimating function is

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

- 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

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

- 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  $K_{\mathcal{V}_n}^*(v, s) = \sum_j K_{\psi_j}^*(v, s)$ , where

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

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

## Saddlepoint approximation (exponential-based)

The saddlepoint density approximation is:

$$g_{\hat{\theta}_n}(s) = \left[ \frac{n}{2\pi \mathcal{K}_{\mathcal{V}_n}^{''}(v_0, s)} \right]^{1/2} e^{\mathcal{K}_{\mathcal{V}_n}^{*}(v_0, s)}, \quad (7)$$

and the saddlepoint  $v_0 = v_0(s)$  solves

$$\mathcal{K}_{\mathcal{V}_n}'(v, s) = 0.$$

### 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  $\sigma^2$ ).

## 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  $\sigma^2$ ). 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  $\sigma^2$ ). 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

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

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

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

where  $v$  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 it

**is asymptotically first order equivalent to the Wald test.**

## 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))

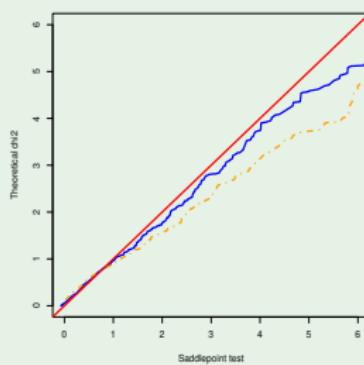
Testing about the long-memory (simple hypothesis) for  $n = 100, 250$ :

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

$$d^0 = 0.1$$

$$d^0 = 0.35$$

Power



# Saddlepoint approximation (exponential-based)

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

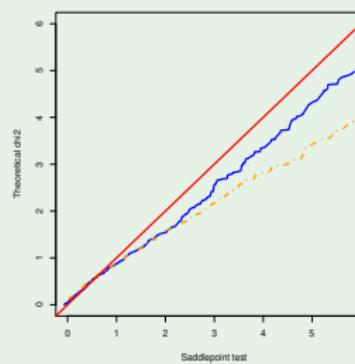
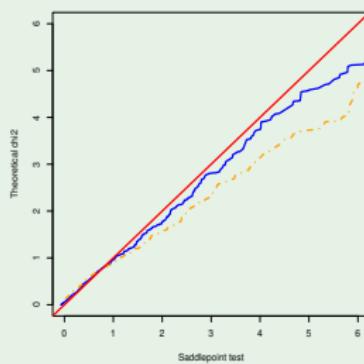
Testing about the long-memory (simple hypothesis) for  $n = 100, 250$ :

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

$$d^0 = 0.1$$

$$d^0 = 0.35$$

Power



# Saddlepoint approximation (exponential-based)

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

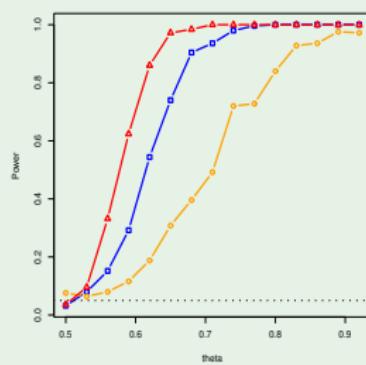
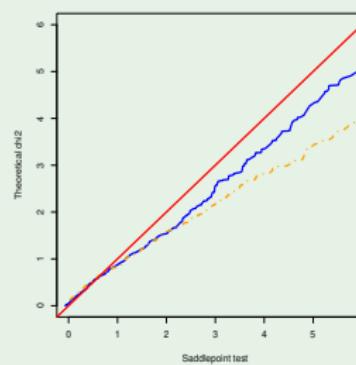
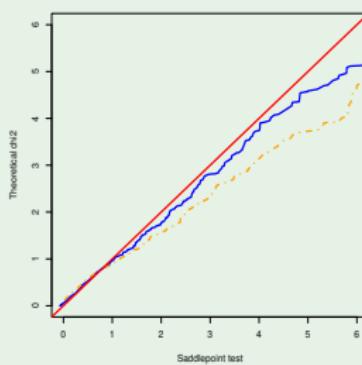
Testing about the long-memory (simple hypothesis) for  $n = 100, 250$ :

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

$$d^0 = 0.1$$

$$d^0 = 0.35$$

Power



## Saddlepoint approximation (empirical version)

### Remark

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

The screenshot shows a Cornell University logo and the text "Cornell University". Below it, the arXiv logo and the URL "arXiv > stat > arXiv:2403.12714" are visible. The title "Statistics > Methodology" is followed by the submission date "Submitted on 19 Mar 2024". The main title of the paper is "On the use of the cumulant generating function for inference on time series". The authors listed are "Alban Moor, Davide La Vecchia, Elvezio Ronchetti". The abstract begins with: "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." The footer includes "Subjects: Methodology (stat.ME); Computation (stat.CO)" and a small logo.

# Saddlepoint approximation (empirical version)

The empirical saddlepoint density approximation is

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

where

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

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

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

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

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

## 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.

## 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.

Indeed, FDEL solves the system of (tilted) estimating equations

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

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

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

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

## Saddlepoint approximation (empirical version)

Now notice that

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

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

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

- 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

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

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

- 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:

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

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 importance:

- (i) it connects our FDES to the FDEL

# 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 importance:

- (i) it connects our FDES to the FDEL
- (ii) it illustrates that the difference between  $\hat{K}$  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:

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

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 importance:

- (i) it connects our FDES to the FDEL
- (ii) it illustrates that the difference between  $\hat{K}$  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 Wald-type (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 Wald-type (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.),

$$W_n(\theta) = n(\hat{\theta}_n - \theta)^T \hat{V}^{-1}(\hat{\theta}_n - \theta).$$

Typically, the distribution of  $W_n$  is approximated by a  $\chi^2$ .

# Saddlepoint approximation (empirical version)

On the practical side: use the empirical saddlepoint under  $\mathcal{H}_0$  to approximate the distribution of Wald-type (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.),

$$W_n(\theta) = n(\hat{\theta}_n - \theta)^T \hat{V}^{-1}(\hat{\theta}_n - \theta).$$

Typically, the distribution of  $W_n$  is approximated by a  $\chi^2$ .

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

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

where  $\tilde{w}(\theta^0)$  is the observed value of the test statistic and

$$\mathcal{B} = \left\{ \theta \in \mathbb{R}^d \mid W_n(\theta) \geq \tilde{w}(\theta^0) \right\}.$$

## Saddlepoint approximation (empirical version)

On the practical side: use the empirical saddlepoint under  $\mathcal{H}_0$  to approximate the distribution of Wald-type (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.),

$$W_n(\theta) = n(\hat{\theta}_n - \theta)^T \hat{V}^{-1}(\hat{\theta}_n - \theta).$$

Typically, the distribution of  $W_n$  is approximated by a  $\chi^2$ .

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

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

where  $\tilde{w}(\theta^0)$  is the observed value of the test statistic and

$$\mathcal{B} = \left\{ \theta \in \mathbb{R}^d \mid W_n(\theta) \geq \tilde{w}(\theta^0) \right\}.$$

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

# Saddlepoint approximation (empirical version)

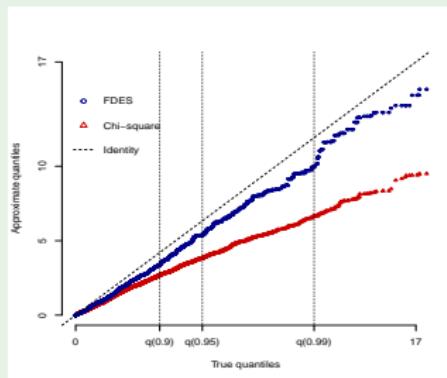
## Example

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

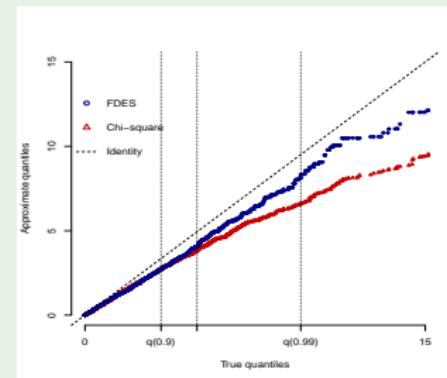
$$\mathcal{H}_0 : d^0 = 0 \text{ vs. } \mathcal{H}_1 : d \neq 0$$

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 ( $\chi_1^2$ )**.

$n = 30$



$n = 250$



# Saddlepoint approximation (empirical version)

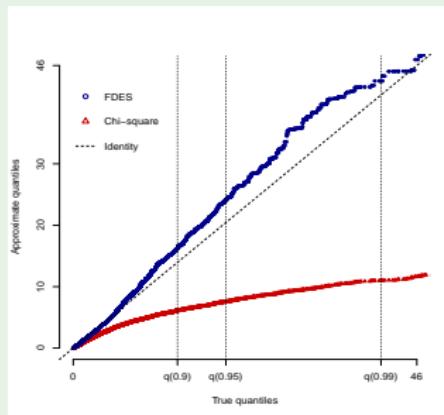
## Example

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

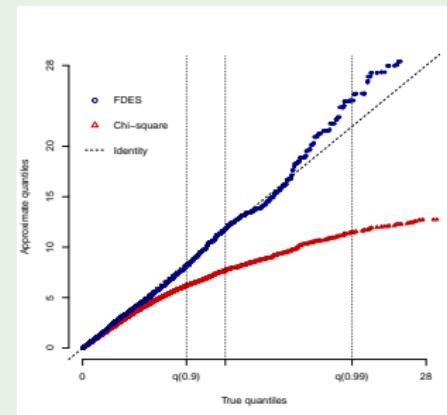
$$\mathcal{H}_0 : \theta = \theta^0 \text{ vs. } \mathcal{H}_1 : \theta \neq \theta^0$$

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 ( $\chi_3^2$ )**.

$n = 100$



$n = 500$



# Saddlepoint approximation (empirical version)

## Remark

- **Testing in the presence of nuisance.** There are cases where only certain components of  $\theta$  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

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

⇒ 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  $\theta$  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](mailto: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:

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

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

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

# 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

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

# 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](#)