Tema 3: Inferencia en procesos de difusión

Francisco de Asís Torres Ruiz

Departamento de Estadística e I.O.

UNIVERSIDAD DE GRANADA

Granada, curso 2023/2024

Índice

1. Introducción
2. Notas sobre la estimación máximo verosímil
3. Inferencia máximo verosímil en procesos de difusión
4. Ejemplos
    - El proceso Wiener
    - El proceso de Ornstein-Uhlenbeck con factores exógenos

Introducción

A continuación abordaremos la estimación máximo verosímil en procesos de difusión mediante muestreo discreto, esto es, supondremos que se dispone de observaciones del proceso en instantes de tiempo $t_1, \ldots, t_n$ en los cuales se observan las variables $X(t_1), \ldots, X(t_n)$, cuyos valores observados constituirán la muestra base del estudio inferencial.

Introducción

A continuación abordaremos la estimación máximo verosímil en procesos de difusión mediante muestreo discreto, esto es, supondremos que se dispone de observaciones del proceso en instantes de tiempo $t_1, \ldots, t_n$ en los cuales se observan las variables $X(t_1), \ldots, X(t_n)$, cuyos valores observados constituirán la muestra base del estudio inferencial.

Para ello será necesario conocer, excepto para valores de parámetros desconocidos, la distribución conjunta de la muestra observada, lo cual conlleva conocer las distribuciones finito-dimensionales del proceso.

Introducción

A continuación abordaremos la estimación máximo verosímil en procesos de difusión mediante muestreo discreto, esto es, supondremos que se dispone de observaciones del proceso en instantes de tiempo $t_1, \ldots, t_n$ en los cuales se observan las variables $X(t_1), \ldots, X(t_n)$, cuyos valores observados constituirán la muestra base del estudio inferencial.

Para ello será necesario conocer, excepto para valores de parámetros desconocidos, la distribución conjunta de la muestra observada, lo cual conlleva conocer las distribuciones finito-dimensionales del proceso.

En nuestro caso, y puesto que los procesos de difusión son procesos de Markov (en realidad, el procedimiento es válido para procesos markovianos en general), la propiedad de Markov permite que a partir de la distribución inicial del proceso y las transiciones se tenga cualquier distribución finito dimensional y, con ello, podamos aplicar la teoría de estimación máximo verosímil.

Estimación máximo verosímil

Este procedimiento fue propuesto y desarrollado por Fisher y es, con mucho, la técnica más popular para calcular estimadores puntuales, amén de estar sustentado por un importante principio de reducción de datos y tener buenas propiedades, sobre todo asintóticas.

El espacio paramétrico es $\Theta = \{1/4, 1/2, 3/4\}$ *y, tras las dos extracciones, el número de bolas blancas extraídas puede ser cero, uno o dos, si bien la probabilidad de cada una de esas posibilidades dependerá del valor de* θ*. En concreto tenemos la siguiente tabla*

		θ		
		$1/4$	$1/2$	$3/4$
x	0	$9/16$	$1/4$	$1/16$
	1	$3/8$	$1/2$	$3/8$
	2	$1/16$	$1/4$	$9/16$

Estimación máximo verosímil

Este procedimiento fue propuesto y desarrollado por Fisher y es, con mucho, la técnica más popular para calcular estimadores puntuales, amén de estar sustentado por un importante principio de reducción de datos y tener buenas propiedades, sobre todo asintóticas. La esencia del método de la máxima verosimilitud radica en que, una vez que una muestra ha sido observada (o el valor concreto de un estadístico muestral) se busca el valor del parámetro que haga que el suceso observado tenga la máxima probabilidad de haber ocurrido.

El espacio paramétrico es $\Theta = \{1/4, 1/2, 3/4\}$ y, tras las dos extracciones, el número de bolas blancas extraídas puede ser cero, uno o dos, si bien la probabilidad de cada una de esas posibilidades dependerá del valor de θ. En concreto tenemos la siguiente tabla

x \ θ	$1/4$	$1/2$	$3/4$
0	$9/16$	$1/4$	$1/16$
1	$3/8$	$1/2$	$3/8$
2	$1/16$	$1/4$	$9/16$

Estimación máximo verosímil

Este procedimiento fue propuesto y desarrollado por Fisher y es, con mucho, la técnica más popular para calcular estimadores puntuales, amén de estar sustentado por un importante principio de reducción de datos y tener buenas propiedades, sobre todo asintóticas. La esencia del método de la máxima verosimilitud radica en que, una vez que una muestra ha sido observada (o el valor concreto de un estadístico muestral) se busca el valor del parámetro que haga que el suceso observado tenga la máxima probabilidad de haber ocurrido. Veamos un ejemplo:

El espacio paramétrico es $\Theta = \{1/4, 1/2, 3/4\}$ y, tras las dos extracciones, el número de bolas blancas extraídas puede ser cero, uno o dos, si bien la probabilidad de cada una de esas posibilidades dependerá del valor de θ. En concreto tenemos la siguiente tabla

		θ		
		$1/4$	$1/2$	$3/4$
x	0	$9/16$	$1/4$	$1/16$
	1	$3/8$	$1/2$	$3/8$
	2	$1/16$	$1/4$	$9/16$

Estimación máximo verosímil

Este procedimiento fue propuesto y desarrollado por Fisher y es, con mucho, la técnica más popular para calcular estimadores puntuales, amén de estar sustentado por un importante principio de reducción de datos y tener buenas propiedades, sobre todo asintóticas. La esencia del método de la máxima verosimilitud radica en que, una vez que una muestra ha sido observada (o el valor concreto de un estadístico muestral) se busca el valor del parámetro que haga que el suceso observado tenga la máxima probabilidad de haber ocurrido. Veamos un ejemplo:

Ejemplo

Una bolsa contiene 4 bolas, de las cuales unas son blancas y otras negras. Al menos hay una de cada color y se desea estimar la proporción de bolas blancas, para lo cual se permite realizar dos extracciones con reemplazamiento.

Estimación máximo verosímil

Este procedimiento fue propuesto y desarrollado por Fisher y es, con mucho, la técnica más popular para calcular estimadores puntuales, amén de estar sustentado por un importante principio de reducción de datos y tener buenas propiedades, sobre todo asintóticas. La esencia del método de la máxima verosimilitud radica en que, una vez que una muestra ha sido observada (o el valor concreto de un estadístico muestral) se busca el valor del parámetro que haga que el suceso observado tenga la máxima probabilidad de haber ocurrido. Veamos un ejemplo:

Ejemplo

Una bolsa contiene 4 bolas, de las cuales unas son blancas y otras negras. Al menos hay una de cada color y se desea estimar la proporción de bolas blancas, para lo cual se permite realizar dos extracciones con reemplazamiento.

El espacio paramétrico es $\Theta = \{1/4, 1/2, 3/4\}$ *y, tras las dos extracciones, el número de bolas blancas extraídas puede ser cero, uno o dos, si bien la probabilidad de cada una de esas posibilidades dependerá del valor de* θ*. En concreto tenemos la siguiente tabla*

Estimación máximo verosímil

Este procedimiento fue propuesto y desarrollado por Fisher y es, con mucho, la técnica más popular para calcular estimadores puntuales, amén de estar sustentado por un importante principio de reducción de datos y tener buenas propiedades, sobre todo asintóticas. La esencia del método de la máxima verosimilitud radica en que, una vez que una muestra ha sido observada (o el valor concreto de un estadístico muestral) se busca el valor del parámetro que haga que el suceso observado tenga la máxima probabilidad de haber ocurrido. Veamos un ejemplo:

Ejemplo

Una bolsa contiene 4 bolas, de las cuales unas son blancas y otras negras. Al menos hay una de cada color y se desea estimar la proporción de bolas blancas, para lo cual se permite realizar dos extracciones con reemplazamiento.

El espacio paramétrico es $\Theta = \{1/4, 1/2, 3/4\}$ *y, tras las dos extracciones, el número de bolas blancas extraídas puede ser cero, uno o dos, si bien la probabilidad de cada una de esas posibilidades dependerá del valor de* θ*. En concreto tenemos la siguiente tabla*

x \ θ	$1/4$	$1/2$	$3/4$
0	$9/16$	$1/4$	$1/16$
1	$3/8$	$1/2$	$3/8$
2	$1/16$	$1/4$	$9/16$

Estimación máximo verosímil

Este procedimiento fue propuesto y desarrollado por Fisher y es, con mucho, la técnica más popular para calcular estimadores puntuales, amén de estar sustentado por un importante principio de reducción de datos y tener buenas propiedades, sobre todo asintóticas. La esencia del método de la máxima verosimilitud radica en que, una vez que una muestra ha sido observada (o el valor concreto de un estadístico muestral) se busca el valor del parámetro que haga que el suceso observado tenga la máxima probabilidad de haber ocurrido. Veamos un ejemplo:

Ejemplo

Una bolsa contiene 4 bolas, de las cuales unas son blancas y otras negras. Al menos hay una de cada color y se desea estimar la proporción de bolas blancas, para lo cual se permite realizar dos extracciones con reemplazamiento.

El espacio paramétrico es $\Theta = \{1/4, 1/2, 3/4\}$ *y, tras las dos extracciones, el número de bolas blancas extraídas puede ser cero, uno o dos, si bien la probabilidad de cada una de esas posibilidades dependerá del valor de* θ*. En concreto tenemos la siguiente tabla*

x \ θ	$1/4$	$1/2$	$3/4$
0	$9/16$	$1/4$	$1/16$
1	$3/8$	$1/2$	$3/8$
2	$1/16$	$1/4$	$9/16$

Si el resultado de las dos extracciones es cero bolas blancas, el valor más verosímil de θ es $1/4$ ya que es el que determina que ese suceso haya ocurrido con mayor probabilidad.

Estimación máximo verosímil

Este procedimiento fue propuesto y desarrollado por Fisher y es, con mucho, la técnica más popular para calcular estimadores puntuales, amén de estar sustentado por un importante principio de reducción de datos y tener buenas propiedades, sobre todo asintóticas. La esencia del método de la máxima verosimilitud radica en que, una vez que una muestra ha sido observada (o el valor concreto de un estadístico muestral) se busca el valor del parámetro que haga que el suceso observado tenga la máxima probabilidad de haber ocurrido. Veamos un ejemplo:

Ejemplo

Una bolsa contiene 4 bolas, de las cuales unas son blancas y otras negras. Al menos hay una de cada color y se desea estimar la proporción de bolas blancas, para lo cual se permite realizar dos extracciones con reemplazamiento.

El espacio paramétrico es $\Theta = \{1/4, 1/2, 3/4\}$ *y, tras las dos extracciones, el número de bolas blancas extraídas puede ser cero, uno o dos, si bien la probabilidad de cada una de esas posibilidades dependerá del valor de* θ*. En concreto tenemos la siguiente tabla*

x \ θ	$1/4$	$1/2$	$3/4$
0	$9/16$	$1/4$	$1/16$
1	$3/8$	$1/2$	$3/8$
2	$1/16$	$1/4$	$9/16$

Si el resultado de las dos extracciones es cero bolas blancas, el valor más verosímil de θ es $1/4$ ya que es el que determina que ese suceso haya ocurrido con mayor probabilidad.

De igual forma, si se obtiene una bola blanca, el valor de θ que otorga mayor probabilidad al suceso observado es $1/2$. Por último, en el caso $x = 2$ bolas blancas, el valor $\theta = 3/4$ es el que hace más verosímil el resultado obtenido.

Estimación máximo verosímil

Este procedimiento fue propuesto y desarrollado por Fisher y es, con mucho, la técnica más popular para calcular estimadores puntuales, amén de estar sustentado por un importante principio de reducción de datos y tener buenas propiedades, sobre todo asintóticas. La esencia del método de la máxima verosimilitud radica en que, una vez que una muestra ha sido observada (o el valor concreto de un estadístico muestral) se busca el valor del parámetro que haga que el suceso observado tenga la máxima probabilidad de haber ocurrido. Veamos un ejemplo:

Ejemplo

Una bolsa contiene 4 bolas, de las cuales unas son blancas y otras negras. Al menos hay una de cada color y se desea estimar la proporción de bolas blancas, para lo cual se permite realizar dos extracciones con reemplazamiento.

El espacio paramétrico es $\Theta = \{1/4, 1/2, 3/4\}$ y, tras las dos extracciones, el número de bolas blancas extraídas puede ser cero, uno o dos, si bien la probabilidad de cada una de esas posibilidades dependerá del valor de θ. En concreto tenemos la siguiente tabla

x \ θ	$1/4$	$1/2$	$3/4$
0	$9/16$	$1/4$	$1/16$
1	$3/8$	$1/2$	$3/8$
2	$1/16$	$1/4$	$9/16$

Si el resultado de las dos extracciones es cero bolas blancas, el valor más verosímil de θ es $1/4$ ya que es el que determina que ese suceso haya ocurrido con mayor probabilidad.

De igual forma, si se obtiene una bola blanca, el valor de θ que otorga mayor probabilidad al suceso observado es $1/2$. Por último, en el caso $x = 2$ bolas blancas, el valor $\theta = 3/4$ es el que hace más verosímil el resultado obtenido.

En definitiva, es razonable emplear como estimador de θ la función

$$\widehat{\theta}(x) = \begin{cases} 1/4 & \text{si } x = 0 \\ 1/2 & \text{si } x = 1 \\ 3/4 & \text{si } x = 2 \end{cases}$$

Estimación máximo verosímil

Este procedimiento fue propuesto y desarrollado por Fisher y es, con mucho, la técnica más popular para calcular estimadores puntuales, amén de estar sustentado por un importante principio de reducción de datos y tener buenas propiedades, sobre todo asintóticas. La esencia del método de la máxima verosimilitud radica en que, una vez que una muestra ha sido observada (o el valor concreto de un estadístico muestral) se busca el valor del parámetro que haga que el suceso observado tenga la máxima probabilidad de haber ocurrido. Veamos un ejemplo:

Ejemplo

Una bolsa contiene 4 bolas, de las cuales unas son blancas y otras negras. Al menos hay una de cada color y se desea estimar la proporción de bolas blancas, para lo cual se permite realizar dos extracciones con reemplazamiento.

El espacio paramétrico es $\Theta = \{1/4, 1/2, 3/4\}$ y, tras las dos extracciones, el número de bolas blancas extraídas puede ser cero, uno o dos, si bien la probabilidad de cada una de esas posibilidades dependerá del valor de θ. En concreto tenemos la siguiente tabla

x \ θ	$1/4$	$1/2$	$3/4$
0	$9/16$	$1/4$	$1/16$
1	$3/8$	$1/2$	$3/8$
2	$1/16$	$1/4$	$9/16$

Si el … que
deterr…

El ejemplo anterior se puede describir mediante una variable binomial $B(1,\theta)$ de la cual se toma una m.a.s. de tamaño dos y se considera el estadístico $T(\mathbf{X}) = X_1 + X_2$, con distribución binomial $B(2,\theta)$, o bien mediante una variable binomial $B(2,\theta)$ de la cual se toma una muestra de tamaño uno.

De igual forma, si se obtiene una bola blanca, el valor de θ que otorga mayor probabilidad al suceso observado es $1/2$. Por último, en el caso $x=2$ bolas blancas, el valor $\theta = 3/4$ es el que hace más verosímil el resultado obtenido.

En definitiva, es razonable emplear como estimador de θ la función

$$\widehat{\theta}(x) = \begin{cases} 1/4 & \text{si } x = 0 \\ 1/2 & \text{si } x = 1 \\ 3/4 & \text{si } x = 2 \end{cases}$$

Estimación máximo verosímil

Este procedimiento fue propuesto y desarrollado por Fisher y es, con mucho, la técnica más popular para calcular estimadores puntuales, amén de estar sustentado por un importante principio de reducción de datos y tener buenas propiedades, sobre todo asintóticas. La esencia del método de la máxima verosimilitud radica en que, una vez que una muestra ha sido observada (o el valor concreto de un estadístico muestral) se busca el valor del parámetro que haga que el suceso observado tenga la máxima probabilidad de haber ocurrido. Veamos un ejemplo:

Ejemplo

Una bolsa contiene 4 bolas, de las cuales unas son blancas y otras negras. Al menos hay una de cada color y se desea estimar la proporción de bolas blancas, para lo cual se permite realizar dos extracciones con reemplazamiento.

El espacio paramétrico es $\Theta = \{1/4, 1/2, 3/4\}$ y, tras las dos extracciones, el número de bolas blancas extraídas puede ser cero, uno o dos, si bien la probabilidad de cada una de esas posibilidades dependerá del valor de θ. En concreto tenemos la siguiente tabla

x \ θ	$1/4$	$1/2$	$3/4$
0	$9/16$	$1/4$	$1/16$
1	$3/8$	$1/2$	$3/8$
2	$1/16$	$1/4$	$9/16$

Si el ... que
deterr...

De ig... o es
$1/2$. F... ido.
En de...

Lo que se ha hecho es buscar, para un valor observado de la muestra, aquel valor del parámetro que hace que la muestra observada tenga la mayor probabilidad de haberse obtenido. Es decir, hemos considerado la función masa de probabilidad (la variable es discreta) de la muestra, hemos fijado un valor muestral y hemos buscado el valor de θ que maximiza dicha función. Obviamente, en este caso el argumento de la función ha sido el parámetro, por lo que no estamos interpretando dicha función como una masa de probabilidad.

$$\widehat{\theta}(x) = \begin{cases} 1/4 & \text{si } x = 0 \\ 1/2 & \text{si } x = 1 \\ 3/4 & \text{si } x = 2 \end{cases}$$

Estimación máximo verosímil

Este procedimiento fue propuesto y desarrollado por Fisher y es, con mucho, la técnica más popular para calcular estimadores puntuales, amén de estar sustentado por un importante principio de reducción de datos y tener buenas propiedades, sobre todo asintóticas. La esencia del método de la máxima verosimilitud radica en que, una vez que una muestra ha sido observada (o el valor concreto de un estadístico muestral) se busca el valor del parámetro que haga que el suceso observado tenga la máxima probabilidad de haber ocurrido. Veamos un ejemplo:

Ejemplo

Una bolsa contiene 4 bolas, de las cuales unas son blancas y otras negras. Al menos hay una de cada color y se desea estimar la proporción de bolas blancas, para lo cual se permite realizar dos extracciones con reemplazamiento.

El espacio paramétrico es $\Theta = \{1/4, 1/2, 3/4\}$ y, tras las dos extracciones, el número de bolas blancas extraídas puede ser cero, uno o dos, si bien la probabilidad de cada una de esas posibilidades dependerá del valor de θ. En concreto tenemos la siguiente tabla

x \ θ	$1/4$	$1/2$	$3/4$
0	$9/16$	$1/4$	$1/16$
1	$3/8$	$1/2$	$3/8$

Comentario

Si $\mathbf{X}$ *es un vector aleatorio discreto, entonces* $L_{\mathbf{x}}(\theta) = \mathrm{P}_\theta(\mathbf{X} = \mathbf{x})$, *por lo que si disponemos de dos puntos del espacio paramétrico,* θ_1 *y* θ_2, *tales que*

$$\mathrm{P}_{\theta_1}(\mathbf{X} = \mathbf{x}) = L_{\mathbf{x}}(\theta_1) > L_{\mathbf{x}}(\theta_2) = \mathrm{P}_{\theta_2}(\mathbf{X} = \mathbf{x}),$$

entonces, dado el valor muestral $\mathbf{x}$ *observado, se puede decir que el valor* θ_1 *tiene más posibilidades de ser el verdadero valor del parámetro que* θ_2. *(Obsérvese que no se ha dicho más probable ya que el parámetro no es una variable aleatoria).*

Si el … que
deterr…

De ig… o es
$1/2$. F… ido.

En de…

$3/4$ si $x = 2$

Estimación máximo verosímil

Este procedimiento fue propuesto y desarrollado por Fisher y es, con mucho, la técnica más popular para calcular estimadores puntuales, amén de estar sustentado por un importante principio de reducción de datos y tener buenas propiedades, sobre todo asintóticas. La esencia del método de la máxima verosimilitud radica en que, una vez que una muestra ha sido observada (o el valor concreto de un estadístico muestral) se busca el valor del parámetro que haga que el suceso observado tenga la máxima probabilidad de haber ocurrido. Veamos un ejemplo:

Ejemplo

Una bolsa contiene 4 bolas, de las cuales unas son blancas y otras negras. Al menos hay una de cada color y se desea estimar la proporción de bolas blancas, para lo cual se permite realizar dos extracciones con reemplazamiento.

El espacio paramétrico es $\Theta = \{1/4, 1/2, 3/4\}$ *y, tras las dos extracciones, el número de bolas blancas extraídas puede ser cero, uno o dos, si bien la probabilidad de cada una de esas posibilidades dependerá del valor de* θ*. En concreto tenemos la siguiente tabla*

θ

x	$1/4$	$1/2$	$3/4$
0	$9/16$	$1/4$	$1/16$
1	$3/8$	$1/2$	$3/8$

Si el ... que
determ...
De ig... o es
$1/2$. F... ido.
En de...

Comentario

Si X *es una variable aleatoria continua unidimensional, entonces (recurriendo al concepto de derivada) para un valor* ϵ *pequeño se tiene que* $\mathrm{P}_\theta(x-\epsilon < X < x+\epsilon)$ *es aproximadamente* $2f_\theta(x) = 2L_x(\theta)$ *(un resultado análogo se tiene en el caso multidimensional con los cambios obvios). Por lo tanto*

$$\frac{\mathrm{P}_{\theta_1}(x-\epsilon < X < x+\epsilon)}{\mathrm{P}_{\theta_2}(x-\epsilon < X < x+\epsilon)} = \frac{L_x(\theta_1)}{L_x(\theta_2)}$$

por lo que la comparación de la función de verosimilitud en θ_1 *y* θ_2 *proporciona una comparación aproximada de la probabilidad del valor muestral* $\mathbf{x}$ *observado.*

Estimación máximo verosímil

En ese sentido, el *Principio de Máxima Verosimilitud* asume que la muestra es representativa de la población y elige como estimación del parámetro aquel valor que maximiza la función de verosimilitud.

Definición

El principio de estimación máximo verosímil consiste en la búsqueda de una función que maximize, para un valor muestral $\mathbf{x}$ *observado, la función de verosimilitud* $L_{\mathbf{x}}(\boldsymbol{\theta})$*; esto es encontrar una función* $\widehat{\boldsymbol{\theta}} : \mathbb{R}^n \to \boldsymbol{\Theta} \subseteq \mathbb{R}^k$ *tal que*

$$L_{\mathbf{x}}(\widehat{\boldsymbol{\theta}}) = \operatorname*{Sup}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} L_{\mathbf{x}}(\theta).$$

En caso de existir, diremos que $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}(\mathbf{X})$ *es un estimador máximo verosímil (EMV) para* $\boldsymbol{\theta}$.

Estimación máximo verosímil

En ese sentido, el *Principio de Máxima Verosimilitud* asume que la muestra es representativa de la población y elige como estimación del parámetro aquel valor que maximiza la función de verosimilitud.

Definición

El principio de estimación máximo verosímil consiste en la búsqueda de una función que maximize, para un valor muestral $\mathbf{x}$ *observado, la función de verosimilitud* $L_{\mathbf{x}}(\boldsymbol{\theta})$*; esto es encontrar una función* $\widehat{\boldsymbol{\theta}} : \mathbb{R}^n \to \boldsymbol{\Theta} \subseteq \mathbb{R}^k$ *tal que*

$$L_{\mathbf{x}}(\widehat{\boldsymbol{\theta}}) = \operatorname*{Sup}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L_{\mathbf{x}}(\theta).$$

En caso de existir, diremos que $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}(\mathbf{X})$ *es un estimador máximo verosímil (EMV) para* $\boldsymbol{\theta}$*.*

Comentario

Puesto que el problema de encontrar el EMV es, casi siempre, analítico, conviene simplificar la situación. Por ello, como la función logaritmo es monótona, suele emplearse en el procedimiento el logaritmo de la función de verosimilitud.

Estimación máximo verosímil

En ese sentido, el *Principio de Máxima Verosimilitud* asume que la muestra es representativa de la población y elige como estimación del parámetro aquel valor que maximiza la función de verosimilitud.

Definición

El principio de estimación máximo verosímil consiste en la búsqueda de una función que maximize, para un valor muestral $\mathbf{x}$ *observado, la función de verosimilitud* $L_{\mathbf{x}}(\boldsymbol{\theta})$*; esto es encontrar una función* $\widehat{\boldsymbol{\theta}} : \mathbb{R}^n \to \boldsymbol{\Theta} \subseteq \mathbb{R}^k$ *tal que*

$$L_{\mathbf{x}}(\widehat{\boldsymbol{\theta}}) = \operatorname*{Sup}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L_{\mathbf{x}}(\theta).$$

En caso de existir, diremos que $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}(\mathbf{X})$ *es un estimador máximo verosímil (EMV) para* $\boldsymbol{\theta}$*.*

Comentario

Puesto que el problema de encontrar el EMV es, casi siempre, analítico, conviene simplificar la situación. Por ello, como la función logaritmo es monótona, suele emplearse en el procedimiento el logaritmo de la función de verosimilitud.

Si $\boldsymbol{\Theta} \subseteq \mathbb{R}^k$ y la verosimilitud es diferenciable en $\boldsymbol{\theta}_i$, $i = 1, \ldots, k$, *posibles candidatos* a configurar el EMV son los valores $\theta_1, \ldots, \theta_k$ solución del sistema de ecuaciones

$$\frac{\partial L_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_i} = 0, \quad i = 1, \ldots, k.$$

Estimación máximo verosímil

En ese sentido, el *Principio de Máxima Verosimilitud* asume que la muestra es representativa de la población y elige como estimación del parámetro aquel valor que maximiza la función de verosimilitud.

Definición

El principio de estimación máximo verosímil consiste en la búsqueda de una función que maximize, para un valor muestral $\mathbf{x}$ *observado, la función de verosimilitud* $L_{\mathbf{x}}(\boldsymbol{\theta})$*; esto es encontrar una función* $\widehat{\boldsymbol{\theta}} : \mathbb{R}^n \to \boldsymbol{\Theta} \subseteq \mathbb{R}^k$ *tal que*

$$L_{\mathbf{x}}(\widehat{\boldsymbol{\theta}}) = \operatorname*{Sup}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L_{\mathbf{x}}(\theta).$$

En caso de existir, diremos que $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}(\mathbf{X})$ *es un estimador máximo verosímil (EMV) para* $\boldsymbol{\theta}$.

Comentario

Puesto que el problema de encontrar el EMV es, casi siempre, analítico, conviene simplificar la situación. Por ello, como la función logaritmo es monótona, suele emplearse en el procedimiento el logaritmo de la función de verosimilitud.

Si $\boldsymbol{\Theta} \subseteq \mathbb{R}^k$ y la verosimilitud es diferenciable en $\boldsymbol{\theta}_i$, $i = 1, \ldots, k$, *posibles candidatos* a configurar el EMV son los valores $\theta_1, \ldots, \theta_k$ solución del sistema de ecuaciones

$$\frac{\partial L_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_i} = 0, \quad i = 1, \ldots, k.$$

Sin embargo, hay que tener precaución ya que esa condición es sólo necesaria para la existencia de una máximo (podría ser un mínimo o un punto de inflexión). Además, aún así, sólo se aseguraría la existencia de un punto extremo local situado en el interior de $\boldsymbol{\Theta}$, pudiendo ocurrir que el máximo se alcanzara en la frontera de $\boldsymbol{\Theta}$ (observemos que lo que se busca en un máximo global de la función de verosimilitud). Por lo tanto, la frontera debe ser examinada por separado.

Estimación máximo verosímil

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)^t$ *una m.a.s. de una población normal* $N_1[\theta, 1]$, $\theta \in \mathbb{R}$. *La función de verosimilitud es*

$$L_{\mathbf{x}}(\theta) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2\right)$$

y su logaritmo $\log\left(L_{\mathbf{x}}(\theta)\right) = -\dfrac{n}{2}\log(2\pi) - \frac{1}{2}\displaystyle\sum_{i=1}^{n}(x_i - \theta)^2$.

Estimación máximo verosímil

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)^t$ *una m.a.s. de una población normal* $N_1[\theta, 1]$, $\theta \in \mathbb{R}$. *La función de verosimilitud es*

$$L_{\mathbf{x}}(\theta) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2\right)$$

y su logaritmo $\log\left(L_{\mathbf{x}}(\theta)\right) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2$. *La ecuación de verosimilitud* $\frac{d\log\left(L_{\mathbf{x}}(\theta)\right)}{d\theta} = 0$ *conduce a* $\sum_{i=1}^{n}(x_i - \theta) = 0$, *cuya solución es* $\widehat{\theta} = \overline{x}$.

Estimación máximo verosímil

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)^t$ *una m.a.s. de una población normal* $N_1[\theta, 1]$, $\theta \in \mathbb{R}$. *La función de verosimilitud es*

$$L_{\mathbf{x}}(\theta) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2\right)$$

y su logaritmo $\log\left(L_{\mathbf{x}}(\theta)\right) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2$. *La ecuación de verosimilitud* $\frac{d\log\left(L_{\mathbf{x}}(\theta)\right)}{d\theta} = 0$ *conduce a* $\sum_{i=1}^{n}(x_i - \theta) = 0$, *cuya solución es* $\widehat{\theta} = \overline{x}$. *Además, es inmediato comprobar que* $\frac{d^2\log\left(L_{\mathbf{x}}(\theta)\right)}{d\theta^2}\,|_{\theta=\overline{x}} < 0$, *por lo que* $\overline{x}$ *es un máximo local.*

Estimación máximo verosímil

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)^t$ *una m.a.s. de una población normal* $N_1[\theta, 1]$, $\theta \in \mathbb{R}$. *La función de verosimilitud es*

$$L_{\mathbf{x}}(\theta) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2\right)$$

y su logaritmo $\log\left(L_{\mathbf{x}}(\theta)\right) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2$. *La ecuación de verosimilitud* $\frac{d\log\left(L_{\mathbf{x}}(\theta)\right)}{d\theta} = 0$ *conduce a* $\sum_{i=1}^{n}(x_i-\theta) = 0$, *cuya solución es* $\widehat{\theta} = \overline{x}$. *Además, es inmediato comprobar que* $\frac{d^2\log\left(L_{\mathbf{x}}(\theta)\right)}{d\theta^2}\,|_{\theta=\overline{x}} < 0$, *por lo que* $\overline{x}$ *es un máximo local.*

Para finalizar, hay que comprobar que ese punto es un máximo global, para lo cual hay que examinar, dada la continuidad de $L_{\mathbf{x}}$ *respecto de* θ, *qué ocurre en* $\pm\infty$.

Estimación máximo verosímil

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)^t$ *una m.a.s. de una población normal* $N_1[\theta, 1]$, $\theta \in \mathbb{R}$. *La función de verosimilitud es*

$$L_{\mathbf{x}}(\theta) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2\right)$$

y su logaritmo $\log\left(L_{\mathbf{x}}(\theta)\right) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2$. *La ecuación de verosimilitud* $\frac{d\log\left(L_{\mathbf{x}}(\theta)\right)}{d\theta} = 0$ *conduce a* $\sum_{i=1}^{n}(x_i - \theta) = 0$, *cuya solución es* $\widehat{\theta} = \overline{x}$. *Además, es inmediato comprobar que* $\frac{d^2\log\left(L_{\mathbf{x}}(\theta)\right)}{d\theta^2}\,|_{\theta=\overline{x}} < 0$, *por lo que* $\overline{x}$ *es un máximo local.*

Para finalizar, hay que comprobar que ese punto es un máximo global, para lo cual hay que examinar, dada la continuidad de $L_{\mathbf{x}}$ *respecto de* θ, *qué ocurre en* $\pm\infty$.

En este caso es inmediato comprobar que $\lim_{\theta\to\pm\infty} L_{\mathbf{x}}(\theta) = 0$, *por lo que ese máximo es global. En consecuencia,* $\widehat{\theta}(\mathbf{X}) = \overline{\mathbf{X}}$ *es el único estimador máximo verosímil de* θ.

Estimación máximo verosímil

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)^t$ *una m.a.s. de una población normal* $N_1[\theta, 1]$, $\theta \in \mathbb{R}$. *La función de verosimilitud es*

$$L_{\mathbf{x}}(\theta) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2\right)$$

y su logaritmo $\log\left(L_{\mathbf{x}}(\theta)\right) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2$. *La ecuación de verosimilitud* $\frac{d\log\left(L_{\mathbf{x}}(\theta)\right)}{d\theta} = 0$ *conduce a* $\sum_{i=1}^{n}(x_i - \theta) = 0$, *cuya solución es* $\widehat{\theta} = \overline{x}$. *Además, es inmediato comprobar que* $\frac{d^2\log\left(L_{\mathbf{x}}(\theta)\right)}{d\theta^2}\Big|_{\theta=\overline{x}} < 0$, *por lo que* $\overline{x}$ *es un máximo local.*

Para finalizar, hay que comprobar que ese punto es un máximo global, para lo cual hay que examinar, dada la continuidad de $L_{\mathbf{x}}$ *respecto de* θ, *qué ocurre en* $\pm\infty$.

En este caso es inmediato comprobar que $\lim_{\theta\to\pm\infty} L_{\mathbf{x}}(\theta) = 0$, *por lo que ese máximo es global. En consecuencia,* $\widehat{\theta}(\mathbf{X}) = \overline{\mathbf{X}}$ *es el único estimador máximo verosímil de* θ.

En ocasiones puede emplearse otro tipo de argumento que permita obtener el valor donde la verosimilitud alcanza su máximo global.

Estimación máximo verosímil

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)^t$ *una m.a.s. de una población normal* $N_1[\theta, 1]$, $\theta \in \mathbb{R}$. *La función de verosimilitud es*

$$L_{\mathbf{x}}(\theta) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2\right)$$

y su logaritmo $\log\left(L_{\mathbf{x}}(\theta)\right) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2$. *La ecuación de verosimilitud* $\frac{d\log\left(L_{\mathbf{x}}(\theta)\right)}{d\theta} = 0$ *conduce a* $\sum_{i=1}^{n}(x_i - \theta) = 0$, *cuya solución es* $\widehat{\theta} = \overline{x}$. *Además, es inmediato comprobar que* $\frac{d^2\log\left(L_{\mathbf{x}}(\theta)\right)}{d\theta^2}\,|_{\theta=\overline{x}} < 0$, *por lo que* $\overline{x}$ *es un máximo local.*

Para finalizar, hay que comprobar que ese punto es un máximo global, para lo cual hay que examinar, dada la continuidad de $L_{\mathbf{x}}$ *respecto de* θ, *qué ocurre en* $\pm\infty$.

En este caso es inmediato comprobar que $\lim_{\theta\to\pm\infty} L_{\mathbf{x}}(\theta) = 0$, *por lo que ese máximo es global. En consecuencia,* $\widehat{\theta}(\mathbf{X}) = \overline{\mathbf{X}}$ *es el único estimador máximo verosímil de* θ.

En ocasiones puede emplearse otro tipo de argumento que permita obtener el valor donde la verosimilitud alcanza su máximo global. En este caso, es conocido que $\sum_{i=1}^{n}(x_i - a)^2 \geq \sum_{i=1}^{n}(x_i - \overline{x})^2$, $\forall a \in \mathbb{R}$, *dándose la igualdad si y sólo si* $a = \overline{x}$. *Por lo tanto,*

$$\exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2\right) \leq \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - \overline{x})^2\right)$$

verificándose la igualdad si y sólo si $\theta = \overline{x}$, *llegándose a la misma conclusión de antes.*

Estimación máximo verosímil

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)'$ *una m.a.s. de una población normal* $N_1[\mu; \sigma^2]$*. La función de verosimilitud es*

$$L_{\mathbf{x}}(\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right)$$

y

Estimación máximo verosímil

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)'$ *una m.a.s. de una población normal* $N_1[\mu; \sigma^2]$. *La función de verosimilitud es*

$$L_{\mathbf{x}}(\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right)$$

y $\log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2.$

Estimación máximo verosímil

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)'$ *una m.a.s. de una población normal* $N_1[\mu; \sigma^2]$. *La función de verosimilitud es*

$$L_{\mathbf{x}}(\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right)$$

y $\log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$. $\log\left(L_{\mathbf{x}}\right)$ *es continua en* μ *y* σ^2, *con derivadas parciales* $\frac{\partial \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \mu} =$

Estimación máximo verosímil

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)'$ *una m.a.s. de una población normal* $N_1[\mu; \sigma^2]$. *La función de verosimilitud es*

$$L_{\mathbf{x}}(\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right)$$

y $\log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$. $\log\left(L_{\mathbf{x}}\right)$ *es continua en* μ *y* σ^2, *con derivadas parciales* $\frac{\partial \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)$ *y*

Estimación máximo verosímil

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)'$ *una m.a.s. de una población normal* $N_1[\mu; \sigma^2]$. *La función de verosimilitud es*

$$L_{\mathbf{x}}(\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right)$$

y $\log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$. $\log\left(L_{\mathbf{x}}\right)$ *es continua en* μ *y* σ^2, *con derivadas parciales* $\frac{\partial \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)$ *y* $\frac{\partial \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \sigma^2} =$

Estimación máximo verosímil

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)'$ *una m.a.s. de una población normal* $N_1[\mu; \sigma^2]$*. La función de verosimilitud es*

$$L_{\mathbf{x}}(\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right)$$

y $\log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$. $\log\left(L_{\mathbf{x}}\right)$ *es continua en* μ *y* σ^2*, con derivadas parciales* $\frac{\partial \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)$ *y* $\frac{\partial \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2$.

Estimación máximo verosímil

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)'$ *una m.a.s. de una población normal* $N_1[\mu; \sigma^2]$. *La función de verosimilitud es*

$$L_{\mathbf{x}}(\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

y $\log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$. $\log\left(L_{\mathbf{x}}\right)$ *es continua en* μ *y* σ^2, *con derivadas parciales* $\frac{\partial \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$ *y* $\frac{\partial \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2$.

Igualando ambas ecuaciones a cero se obtiene $\widehat{\mu} = \overline{\mathbf{x}}$ *y* $\widehat{\sigma^2} = s^2_{\mathbf{x}}$.

Estimación máximo verosímil

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)'$ *una m.a.s. de una población normal* $N_1[\mu; \sigma^2]$. *La función de verosimilitud es*

$$L_{\mathbf{x}}(\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right)$$

y $\log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$. $\log\left(L_{\mathbf{x}}\right)$ *es continua en* μ *y* σ^2, *con derivadas parciales* $\frac{\partial \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)$ *y* $\frac{\partial \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2$.

Igualando ambas ecuaciones a cero se obtiene $\widehat{\mu} = \overline{\mathbf{x}}$ *y* $\widehat{\sigma^2} = s^2_{\mathbf{x}}$. *Ese es el único punto crítico y comprobamos que es un máximo local. En efecto, las derivadas parciales segundas son*

$$\frac{\partial^2 \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \mu^2} = -\frac{n}{\sigma^2},$$

Estimación máximo verosímil

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)'$ *una m.a.s. de una población normal* $N_1[\mu; \sigma^2]$. *La función de verosimilitud es*

$$L_{\mathbf{x}}(\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right)$$

y $\log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$. $\log\left(L_{\mathbf{x}}\right)$ *es continua en* μ *y* σ^2, *con derivadas parciales* $\frac{\partial \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)$ *y* $\frac{\partial \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2$.

Igualando ambas ecuaciones a cero se obtiene $\widehat{\mu} = \overline{\mathbf{x}}$ *y* $\widehat{\sigma^2} = s^2_{\mathbf{x}}$. *Ese es el único punto crítico y comprobamos que es un máximo local. En efecto, las derivadas parciales segundas son*

$$\frac{\partial^2 \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \mu^2} = -\frac{n}{\sigma^2}, \quad \frac{\partial^2 \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial (\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}(x_i - \mu)^2 \text{ y}$$

Estimación máximo verosímil

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)'$ *una m.a.s. de una población normal* $N_1[\mu; \sigma^2]$*. La función de verosimilitud es*

$$L_{\mathbf{x}}(\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right)$$

y $\log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$. $\log\left(L_{\mathbf{x}}\right)$ *es continua en* μ *y* σ^2*, con derivadas parciales* $\frac{\partial \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)$ *y* $\frac{\partial \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2$.

Igualando ambas ecuaciones a cero se obtiene $\widehat{\mu} = \overline{\mathbf{x}}$ *y* $\widehat{\sigma^2} = s^2_{\mathbf{x}}$*. Ese es el único punto crítico y comprobamos que es un máximo local. En efecto, las derivadas parciales segundas son*

$$\frac{\partial^2 \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \mu^2} = -\frac{n}{\sigma^2},\quad \frac{\partial^2 \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial (\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}(x_i - \mu)^2 \text{ y } \frac{\partial^2 \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \mu \partial \sigma^2} = -\frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i - \mu)$$

Estimación máximo verosímil

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)'$ *una m.a.s. de una población normal* $N_1[\mu; \sigma^2]$. *La función de verosimilitud es*

$$L_{\mathbf{x}}(\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right)$$

y $\log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$. $\log\left(L_{\mathbf{x}}\right)$ *es continua en* μ *y* σ^2, *con derivadas parciales* $\frac{\partial \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)$ *y* $\frac{\partial \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2$.

Igualando ambas ecuaciones a cero se obtiene $\widehat{\mu} = \overline{\mathbf{x}}$ *y* $\widehat{\sigma^2} = s^2_{\mathbf{x}}$. *Ese es el único punto crítico y comprobamos que es un máximo local. En efecto, las derivadas parciales segundas son*

$$\frac{\partial^2 \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \mu^2} = -\frac{n}{\sigma^2},\ \frac{\partial^2 \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial (\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}(x_i - \mu)^2 \text{ y } \frac{\partial^2 \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \mu \partial \sigma^2} = -\frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i - \mu)$$

de donde la matriz Hessiana evaluada en $(\mu, \sigma^2)^t = (\overline{\mathbf{x}}, s^2_{\mathbf{x}})^t$ *es*

$$H(\overline{\mathbf{x}}, s^2_{\mathbf{x}}) = \begin{pmatrix} -\frac{n}{s^2_{\mathbf{x}}} & 0 \\ 0 & -\frac{n}{2(s^2_{\mathbf{x}})^2} \end{pmatrix}$$

Estimación máximo verosímil

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)'$ *una m.a.s. de una población normal* $N_1[\mu; \sigma^2]$*. La función de verosimilitud es*

$$L_{\mathbf{x}}(\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

y $\log\left(L_{\mathbf{x}}(\mu,\sigma^2)\right) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$. $\log\left(L_{\mathbf{x}}\right)$ *es continua en* μ *y* σ^2*, con derivadas parciales* $\frac{\partial \log\left(L_{\mathbf{x}}(\mu,\sigma^2)\right)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$ *y* $\frac{\partial \log\left(L_{\mathbf{x}}(\mu,\sigma^2)\right)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2$.

Igualando ambas ecuaciones a cero se obtiene $\widehat{\mu} = \overline{\mathbf{x}}$ *y* $\widehat{\sigma^2} = s^2_{\mathbf{x}}$*. Ese es el único punto crítico y comprobamos que es un máximo local. En efecto, las derivadas parciales segundas son*

$$\frac{\partial^2 \log\left(L_{\mathbf{x}}(\mu,\sigma^2)\right)}{\partial \mu^2} = -\frac{n}{\sigma^2},\ \frac{\partial^2 \log\left(L_{\mathbf{x}}(\mu,\sigma^2)\right)}{\partial (\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}(x_i-\mu)^2 \text{ y } \frac{\partial^2 \log\left(L_{\mathbf{x}}(\mu,\sigma^2)\right)}{\partial\mu\partial\sigma^2} = -\frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu)$$

de donde la matriz Hessiana evaluada en $(\mu,\sigma^2)^t = (\overline{\mathbf{x}}, s^2_{\mathbf{x}})^t$ *es*

$$H(\overline{\mathbf{x}}, s^2_{\mathbf{x}}) = \begin{pmatrix} -\frac{n}{s^2_{\mathbf{x}}} & 0 \\ 0 & -\frac{n}{2(s^2_{\mathbf{x}})^2} \end{pmatrix}$$

que se puede comprobar es definida negativa.

Estimación máximo verosímil

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)'$ *una m.a.s. de una población normal* $N_1[\mu; \sigma^2]$. *La función de verosimilitud es*

$$L_{\mathbf{x}}(\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right)$$

y $\log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$. $\log\left(L_{\mathbf{x}}\right)$ *es continua en* μ *y* σ^2, *con derivadas parciales* $\frac{\partial \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)$ *y* $\frac{\partial \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2$.

Igualando ambas ecuaciones a cero se obtiene $\widehat{\mu} = \overline{\mathbf{x}}$ *y* $\widehat{\sigma^2} = s^2_{\mathbf{x}}$. *Ese es el único punto crítico y comprobamos que es un máximo local. En efecto, las derivadas parciales segundas son*

$$\frac{\partial^2 \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \mu^2} = -\frac{n}{\sigma^2}, \quad \frac{\partial^2 \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial (\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}(x_i - \mu)^2 \text{ y } \frac{\partial^2 \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \mu \partial \sigma^2} = -\frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i - \mu)$$

de donde la matriz Hessiana evaluada en $(\mu, \sigma^2)^t = (\overline{\mathbf{x}}, s^2_{\mathbf{x}})^t$ *es*

$$H(\overline{\mathbf{x}}, s^2_{\mathbf{x}}) = \begin{pmatrix} -\frac{n}{s^2_{\mathbf{x}}} & 0 \\ 0 & -\frac{n}{2(s^2_{\mathbf{x}})^2} \end{pmatrix}$$

que se puede comprobar es definida negativa.

No obstante, esto demuestra que es un máximo local, pero no global. Ello habría que hacerlo tomando límites de la función de verosimilitud en la frontera de $\Theta = \mathbb{R} \times \mathbb{R}^+$, *pudiéndose comprobar que esos límites son cero.*

Estimación máximo verosímil

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)'$ *una m.a.s. de una población normal* $N_1[\mu; \sigma^2]$. *La función de verosimilitud es*

$$L_{\mathbf{x}}(\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

y $\log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$. $\log\left(L_{\mathbf{x}}\right)$ *es continua en* μ *y* σ^2, *con derivadas parciales* $\frac{\partial \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$ *y* $\frac{\partial \log\left(L_{\mathbf{x}}(\mu, \sigma^2)\right)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2$.

Igualando ambas ecuaciones a cero se obtiene $\widehat{\mu} = \overline{\mathbf{x}}$ *y* $\widehat{\sigma^2} = s^2$. *Ese es el único punto crítico y comprobamos que es un*

No obstante, en este caso él problema se puede reducir a un problema unidimensional ya que $L_{\mathbf{x}}(\overline{\mathbf{x}}, \sigma^2) \geq L_{\mathbf{x}}(\mu, \sigma^2)$, $\forall \sigma^2 > 0$ puesto que $\sum_{i=1}^{n}(x_i-\overline{\mathbf{x}})^2 \leq \sum_{i=1}^{n}(x_i-\mu)^2$, $\forall \mu \in \mathbb{R}$ y dándose la igualdad sólo para $\mu = \overline{\mathbf{x}}$. En tal caso sólo habría que maximizar, en σ^2, la función $L_{\mathbf{x}}(\overline{\mathbf{x}}, \sigma^2)$.

∂^2 lo ... $\mu)$

de donde la matriz Hessiana evaluada en $(\mu, \sigma^2)^t = (\overline{\mathbf{x}}, s^2_{\mathbf{x}})^t$ *es*

$$H(\overline{\mathbf{x}}, s^2_{\mathbf{x}}) = \begin{pmatrix} -\frac{n}{s^2_{\mathbf{x}}} & 0 \\ 0 & -\frac{n}{2(s^2_{\mathbf{x}})^2} \end{pmatrix}$$

que se puede comprobar es definida negativa.

No obstante, esto demuestra que es un máximo local, pero no global. Ello habría que hacerlo tomando límites de la función de verosimilitud en la frontera de $\Theta = \mathbb{R} \times \mathbb{R}^+$, *pudiéndose comprobar que esos límites son cero.*

Estimación máximo verosímil

Sea ahora ψ una función paramétrica y nos planteamos cuál será, de existir, su estimador máximo verosímil.

Observemos que $\mathrm{P}_\theta(X \leq t) = F_\theta(t) = 1 - e^{-\theta\, t}$, *por lo que se trata de calcular el EMV para una función paramétrica* $\psi(\theta) = 1 - e^{-\theta\, t}$. *Calculemos en primer lugar el EMV de* θ.

La verosimilitud es $L_{\mathbf{x}}(\theta) = \theta^n \exp\left(-\theta \sum_{i=1}^{n} x_i\right)$ *y su logaritmo* $\log\left(L_{\mathbf{x}}(\theta)\right) = n\log(\theta) - \theta \sum_{i=1}^{n} x_i$ *es continua en* θ.

La derivada de $\log\left(L_{\mathbf{x}}\right)$ *es* $\dfrac{\partial \log\left(L_{\mathbf{x}}(\theta)\right)}{\partial \theta} = \dfrac{n}{\theta} - \sum_{i=1}^{n} x_i$, *por lo que la ecuación de verosimilitud* $\dfrac{\partial \log\left(L_{\mathbf{x}}(\theta)\right)}{\partial \theta} = 0$ *conduce a la solución* $\widehat{\theta} = \dfrac{1}{\overline{\mathbf{x}}}$. *Veamos que es un máximo.*

En primer lugar observemos que $\dfrac{\partial^2 \log\left(L_{\mathbf{x}}(\theta)\right)}{\partial \theta^2} = -\dfrac{n}{\theta^2}$ *por lo que* $\dfrac{\partial^2 \log\left(L_{\mathbf{x}}(\theta)\right)}{\partial \theta^2}\Big|_{\theta=\overline{\mathbf{x}}} = -n\overline{\mathbf{x}}^2 < 0$ *y con ello* $\widehat{\theta}$ *es un máximo local.*

Para comprobar que es un máximo global, es inmediato verificar que $\lim_{\theta\to 0} L_{\mathbf{x}}(\theta) = \lim_{\theta\to+\infty} L_{\mathbf{x}}(\theta) = 0$, *con lo que ese punto es un máximo global y con ello* $\widehat{\theta}(\mathbf{X}) = 1/\overline{\mathbf{X}}$. *En virtud del Teorema de Zehna,* $\widehat{\psi(\theta)} = 1 - e^{-t/\overline{\mathbf{X}}}$.

Estimación máximo verosímil

Sea ahora ψ una función paramétrica y nos planteamos cuál será, de existir, su estimador máximo verosímil.

Teorema

(de invarianza de Zehna). Si $\widehat{\theta}$ es un estimador máximo verosímil de θ, con $\theta \in \Theta \subseteq \mathbb{R}^k$ y $\psi : \Theta \to \Lambda$ es una función paramétrica con $\Lambda \subseteq \mathbb{R}^p$, entonces $\psi(\widehat{\theta})$ es un estimador máximo verosímil de $\psi(\theta)$.

Observemos que $\mathrm{P}_\theta(X \leq t) = F_\theta(t) = 1 - e^{-\theta t}$, por lo que se trata de calcular el EMV para una función paramétrica $\psi(\theta) = 1 - e^{-\theta t}$. Calculemos en primer lugar el EMV de θ.

La verosimilitud es $L_{\mathbf{x}}(\theta) = \theta^n \exp\left(-\theta \sum_{i=1}^{n} x_i\right)$ y su logaritmo $\log\left(L_{\mathbf{x}}(\theta)\right) = n\log(\theta) - \theta \sum_{i=1}^{n} x_i$ es continua en θ.

La derivada de $\log\left(L_{\mathbf{x}}\right)$ es $\dfrac{\partial \log\left(L_{\mathbf{x}}(\theta)\right)}{\partial\theta} = \dfrac{n}{\theta} - \sum_{i=1}^{n} x_i$, por lo que la ecuación de verosimilitud $\dfrac{\partial \log\left(L_{\mathbf{x}}(\theta)\right)}{\partial\theta} = 0$ conduce a la solución $\widehat{\theta} = \dfrac{1}{\overline{\mathrm{x}}}$. Veamos que es un máximo.

En primer lugar observemos que $\dfrac{\partial^2 \log\left(L_{\mathbf{x}}(\theta)\right)}{\partial\theta^2} = -\dfrac{n}{\theta^2}$ por lo que $\dfrac{\partial^2 \log\left(L_{\mathbf{x}}(\theta)\right)}{\partial\theta^2}\Big|_{\theta=\overline{\mathrm{x}}} = -n\overline{\mathrm{x}}^2 < 0$ y con ello $\widehat{\theta}$ es un máximo local.

Para comprobar que es un máximo global, es inmediato verificar que $\lim\limits_{\theta\to 0} L_{\mathbf{x}}(\theta) = \lim\limits_{\theta\to+\infty} L_{\mathbf{x}}(\theta) = 0$, con lo que ese punto es un máximo global y con ello $\widehat{\theta}(\mathbf{X}) = 1/\overline{\mathbf{X}}$. En virtud del Teorema de Zehna, $\widehat{\psi(\theta)} = 1 - e^{-t/\overline{\mathbf{X}}}$.

Estimación máximo verosímil

Sea ahora ψ una función paramétrica y nos planteamos cuál será, de existir, su estimador máximo verosímil.

Teorema

(de invarianza de Zehna). Si $\widehat{\theta}$ es un estimador máximo verosímil de θ, con $\theta \in \Theta \subseteq \mathbb{R}^k$ y $\psi : \Theta \to \Lambda$ es una función paramétrica con $\Lambda \subseteq \mathbb{R}^p$, entonces $\psi(\widehat{\theta})$ es un estimador máximo verosímil de $\psi(\theta)$.

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)^t$ una m.a.s. de una variable aleatoria con función de densidad $f_\theta(x) = \theta e^{-\theta x}$, $x > 0$, $\theta > 0$. Calcular el EMV para $\mathrm{P}_\theta(X \leq t)$.

Estimación máximo verosímil

Sea ahora ψ una función paramétrica y nos planteamos cuál será, de existir, su estimador máximo verosímil.

Teorema

(de invarianza de Zehna). Si $\widehat{\theta}$ es un estimador máximo verosímil de θ, con $\theta \in \Theta \subseteq \mathbb{R}^k$ y $\psi : \Theta \to \Lambda$ es una función paramétrica con $\Lambda \subseteq \mathbb{R}^p$, entonces $\psi(\widehat{\theta})$ es un estimador máximo verosímil de $\psi(\theta)$.

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)^t$ una m.a.s. de una variable aleatoria con función de densidad $f_\theta(x) = \theta e^{-\theta x}$, $x > 0$, $\theta > 0$. Calcular el EMV para $\mathrm{P}_\theta(X \leq t)$.

Observemos que $\mathrm{P}_\theta(X \leq t) = F_\theta(t) = 1 - e^{-\theta\, t}$, por lo que se trata de calcular el EMV para una función paramétrica $\psi(\theta) = 1 - e^{-\theta\, t}$. Calculemos en primer lugar el EMV de θ.

Estimación máximo verosímil

Sea ahora ψ una función paramétrica y nos planteamos cuál será, de existir, su estimador máximo verosímil.

Teorema

(de invarianza de Zehna). Si $\widehat{\theta}$ es un estimador máximo verosímil de θ, con $\theta \in \Theta \subseteq \mathbb{R}^k$ y $\psi : \Theta \to \Lambda$ es una función paramétrica con $\Lambda \subseteq \mathbb{R}^p$, entonces $\psi(\widehat{\theta})$ es un estimador máximo verosímil de $\psi(\theta)$.

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)^t$ una m.a.s. de una variable aleatoria con función de densidad $f_\theta(x) = \theta e^{-\theta x}$, $x > 0$, $\theta > 0$. Calcular el EMV para $\mathrm{P}_\theta(X \leq t)$.

Observemos que $\mathrm{P}_\theta(X \leq t) = F_\theta(t) = 1 - e^{-\theta\, t}$, por lo que se trata de calcular el EMV para una función paramétrica $\psi(\theta) = 1 - e^{-\theta\, t}$. Calculemos en primer lugar el EMV de θ.

La verosimilitud es $L_{\mathbf{x}}(\theta) = \theta^n \exp\left(-\theta \sum_{i=1}^n x_i\right)$ y su logaritmo $\log\left(L_{\mathbf{x}}(\theta)\right) = n\log(\theta) - \theta\sum_{i=1}^n x_i$ es continua en θ.

Estimación máximo verosímil

Sea ahora ψ una función paramétrica y nos planteamos cuál será, de existir, su estimador máximo verosímil.

Teorema

(de invarianza de Zehna). Si $\widehat{\theta}$ es un estimador máximo verosímil de θ, con $\theta \in \Theta \subseteq \mathbb{R}^k$ y $\psi : \Theta \to \Lambda$ es una función paramétrica con $\Lambda \subseteq \mathbb{R}^p$, entonces $\psi(\widehat{\theta})$ es un estimador máximo verosímil de $\psi(\theta)$.

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)^t$ una m.a.s. de una variable aleatoria con función de densidad $f_\theta(x) = \theta e^{-\theta x}$, $x > 0$, $\theta > 0$. Calcular el EMV para $\mathrm{P}_\theta(X \leq t)$.

Observemos que $\mathrm{P}_\theta(X \leq t) = F_\theta(t) = 1 - e^{-\theta\, t}$, por lo que se trata de calcular el EMV para una función paramétrica $\psi(\theta) = 1 - e^{-\theta\, t}$. Calculemos en primer lugar el EMV de θ.

La verosimilitud es $L_{\mathbf{x}}(\theta) = \theta^n \exp\left(-\theta \sum_{i=1}^{n} x_i\right)$ y su logaritmo $\log\left(L_{\mathbf{x}}(\theta)\right) = n\log(\theta) - \theta \sum_{i=1}^{n} x_i$ es continua en θ.

La derivada de $\log\left(L_{\mathbf{x}}\right)$ es $\dfrac{\partial \log\left(L_{\mathbf{x}}(\theta)\right)}{\partial \theta} = \dfrac{n}{\theta} - \sum_{i=1}^{n} x_i$, por lo que la ecuación de verosimilitud $\dfrac{\partial \log\left(L_{\mathbf{x}}(\theta)\right)}{\partial \theta} = 0$ conduce a la solución $\widehat{\theta} = \dfrac{1}{\overline{\mathbf{x}}}$. Veamos que es un máximo.

Estimación máximo verosímil

Sea ahora ψ una función paramétrica y nos planteamos cuál será, de existir, su estimador máximo verosímil.

Teorema

(de invarianza de Zehna). Si $\widehat{\theta}$ es un estimador máximo verosímil de θ, con $\theta \in \Theta \subseteq \mathbb{R}^k$ y $\psi : \Theta \to \Lambda$ es una función paramétrica con $\Lambda \subseteq \mathbb{R}^p$, entonces $\psi(\widehat{\theta})$ es un estimador máximo verosímil de $\psi(\theta)$.

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)^t$ una m.a.s. de una variable aleatoria con función de densidad $f_\theta(x) = \theta e^{-\theta x}$, $x > 0$, $\theta > 0$. Calcular el EMV para $\mathrm{P}_\theta(X \leq t)$.

Observemos que $\mathrm{P}_\theta(X \leq t) = F_\theta(t) = 1 - e^{-\theta\, t}$, por lo que se trata de calcular el EMV para una función paramétrica $\psi(\theta) = 1 - e^{-\theta\, t}$. Calculemos en primer lugar el EMV de θ.

La verosimilitud es $L_{\mathbf{x}}(\theta) = \theta^n \exp\left(-\theta \sum_{i=1}^n x_i\right)$ y su logaritmo $\log\left(L_{\mathbf{x}}(\theta)\right) = n\log(\theta) - \theta\sum_{i=1}^n x_i$ es continua en θ.

La derivada de $\log\left(L_{\mathbf{x}}\right)$ es $\dfrac{\partial \log\left(L_{\mathbf{x}}(\theta)\right)}{\partial\theta} = \dfrac{n}{\theta} - \sum_{i=1}^n x_i$, por lo que la ecuación de verosimilitud $\dfrac{\partial \log\left(L_{\mathbf{x}}(\theta)\right)}{\partial\theta} = 0$ conduce a la solución $\widehat{\theta} = \dfrac{1}{\overline{\mathbf{x}}}$. Veamos que es un máximo.

En primer lugar observemos que $\dfrac{\partial^2 \log\left(L_{\mathbf{x}}(\theta)\right)}{\partial\theta^2} = -\dfrac{n}{\theta^2}$ por lo que $\dfrac{\partial^2 \log\left(L_{\mathbf{x}}(\theta)\right)}{\partial\theta^2}\Big|_{\theta=\overline{\mathbf{x}}} = -n\overline{\mathbf{x}}^2 < 0$ y con ello $\widehat{\theta}$ es un máximo local.

Estimación máximo verosímil

Sea ahora ψ una función paramétrica y nos planteamos cuál será, de existir, su estimador máximo verosímil.

Teorema

(de invarianza de Zehna). Si $\widehat{\theta}$ es un estimador máximo verosímil de θ, con $\theta \in \Theta \subseteq \mathbb{R}^k$ y $\psi : \Theta \to \Lambda$ es una función paramétrica con $\Lambda \subseteq \mathbb{R}^p$, entonces $\psi(\widehat{\theta})$ es un estimador máximo verosímil de $\psi(\theta)$.

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)^t$ una m.a.s. de una variable aleatoria con función de densidad $f_\theta(x) = \theta e^{-\theta x}$, $x > 0$, $\theta > 0$. Calcular el EMV para $\mathrm{P}_\theta(X \leq t)$.

Observemos que $\mathrm{P}_\theta(X \leq t) = F_\theta(t) = 1 - e^{-\theta\, t}$, por lo que se trata de calcular el EMV para una función paramétrica $\psi(\theta) = 1 - e^{-\theta\, t}$. Calculemos en primer lugar el EMV de θ.

La verosimilitud es $L_{\mathbf{x}}(\theta) = \theta^n \exp\left(-\theta \sum_{i=1}^n x_i\right)$ y su logaritmo $\log\left(L_{\mathbf{x}}(\theta)\right) = n\log(\theta) - \theta \sum_{i=1}^n x_i$ es continua en θ.

La derivada de $\log\left(L_{\mathbf{x}}\right)$ es $\dfrac{\partial \log\left(L_{\mathbf{x}}(\theta)\right)}{\partial\theta} = \dfrac{n}{\theta} - \sum_{i=1}^n x_i$, por lo que la ecuación de verosimilitud $\dfrac{\partial \log\left(L_{\mathbf{x}}(\theta)\right)}{\partial\theta} = 0$ conduce a la solución $\widehat{\theta} = \dfrac{1}{\overline{\mathbf{x}}}$. Veamos que es un máximo.

En primer lugar observemos que $\dfrac{\partial^2 \log\left(L_{\mathbf{x}}(\theta)\right)}{\partial\theta^2} = -\dfrac{n}{\theta^2}$ por lo que $\dfrac{\partial^2 \log\left(L_{\mathbf{x}}(\theta)\right)}{\partial\theta^2}\,|_{\theta=\overline{\mathbf{x}}} = -n\overline{\mathbf{x}}^2 < 0$ y con ello $\widehat{\theta}$ es un máximo local.

Para comprobar que es un máximo global, es inmediato verificar que $\lim\limits_{\theta\to 0} L_{\mathbf{x}}(\theta) = \lim\limits_{\theta\to+\infty} L_{\mathbf{x}}(\theta) = 0$, con lo que ese punto es un máximo global y con ello $\widehat{\theta}(\mathbf{X}) = 1/\overline{\mathbf{X}}$.

Estimación máximo verosímil

Sea ahora ψ una función paramétrica y nos planteamos cuál será, de existir, su estimador máximo verosímil.

Teorema

(de invarianza de Zehna). Si $\widehat{\theta}$ es un estimador máximo verosímil de θ, con $\theta \in \Theta \subseteq \mathbb{R}^k$ y $\psi : \Theta \to \Lambda$ es una función paramétrica con $\Lambda \subseteq \mathbb{R}^p$, entonces $\psi(\widehat{\theta})$ es un estimador máximo verosímil de $\psi(\theta)$.

Ejemplo

Sea $\mathbf{X} = (X_1, \ldots, X_n)^t$ una m.a.s. de una variable aleatoria con función de densidad $f_\theta(x) = \theta e^{-\theta x}$, $x > 0$, $\theta > 0$. Calcular el EMV para $\mathrm{P}_\theta(X \leq t)$.

Observemos que $\mathrm{P}_\theta(X \leq t) = F_\theta(t) = 1 - e^{-\theta\, t}$, por lo que se trata de calcular el EMV para una función paramétrica $\psi(\theta) = 1 - e^{-\theta\, t}$. Calculemos en primer lugar el EMV de θ.

La verosimilitud es $L_{\mathbf{x}}(\theta) = \theta^n \exp\left(-\theta \sum_{i=1}^n x_i\right)$ y su logaritmo $\log\left(L_{\mathbf{x}}(\theta)\right) = n\log(\theta) - \theta \sum_{i=1}^n x_i$ es continua en θ.

La derivada de $\log\left(L_{\mathbf{x}}\right)$ es $\dfrac{\partial \log\left(L_{\mathbf{x}}(\theta)\right)}{\partial \theta} = \dfrac{n}{\theta} - \sum_{i=1}^n x_i$, por lo que la ecuación de verosimilitud $\dfrac{\partial \log\left(L_{\mathbf{x}}(\theta)\right)}{\partial \theta} = 0$ conduce a la solución $\widehat{\theta} = \dfrac{1}{\overline{\mathbf{x}}}$. Veamos que es un máximo.

En primer lugar observemos que $\dfrac{\partial^2 \log\left(L_{\mathbf{x}}(\theta)\right)}{\partial \theta^2} = -\dfrac{n}{\theta^2}$ por lo que $\dfrac{\partial^2 \log\left(L_{\mathbf{x}}(\theta)\right)}{\partial \theta^2}\Big|_{\theta = \overline{\mathbf{x}}} = -n\overline{\mathbf{x}}^2 < 0$ y con ello $\widehat{\theta}$ es un máximo local.

Para comprobar que es un máximo global, es inmediato verificar que $\lim\limits_{\theta \to 0} L_{\mathbf{x}}(\theta) = \lim\limits_{\theta \to +\infty} L_{\mathbf{x}}(\theta) = 0$, con lo que ese punto es un máximo global y con ello $\widehat{\theta}(\mathbf{X}) = 1/\overline{\mathbf{X}}$. En virtud del Teorema de Zehna, $\widehat{\psi(\theta)} = 1 - e^{-t/\overline{\mathbf{X}}}$.

Estimación máximo verosímil

Concluimos comentando cuál es la distribución asintótica del EMV. El siguiente teorema lo expresaremos supuesto que la familia de distribuciones es continua, si bien es cierto en el caso discreto con los cambios obvios.

Estimación máximo verosímil

Concluimos comentando cuál es la distribución asintótica del EMV. El siguiente teorema lo expresaremos supuesto que la familia de distribuciones es continua, si bien es cierto en el caso discreto con los cambios obvios.

Teorema

Sea $\{f_\theta : \theta \in \Theta\}$ una familia paramétrica de densidades con $\Theta \subseteq \mathbb{R}$. Sea una m.a.s. extraída de una población con densidad en dicha familia y supongamos que se verifican las siguientes condiciones:

- *Existen las parciales* $\dfrac{\partial^i \log\left(f_\theta(x)\right)}{\partial\theta^i}$, $i = 1, 2, 3$, $\forall\theta \in A$, *siendo A un abierto incluido en Θ.*
- *Existe una función M integrable tal que* $\left|\dfrac{\partial^3 \log\left(f_\theta(x)\right)}{\partial\theta^3}\right| < M(x)$ *con* $\mathrm{E}_\theta[M(X)] < K(\theta)$, $\forall\theta \in A$.
- *Existe al menos un valor $\theta_0 \in A$ verificando*
  - $\mathrm{E}_{\theta_0}\left[\dfrac{\partial \log\left(f_\theta(X)\right)}{\partial\theta}\right] = 0$, $\mathrm{E}_{\theta_0}\left[\dfrac{1}{f_\theta(X)}\dfrac{\partial^2 f_\theta(X)}{\partial\theta^2}\right] = 0$
  - $0 < I(\theta_0) < +\infty$, *con $I(\theta)$ la función de información de Fisher basado en una muestra de tamaño uno.*

Entonces cualquier sucesión débilmente consistente de raíces de la ecuación de verosimilitud verifica la convergencia en ley $\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{L} Z$, *donde* $Z \rightsquigarrow N_1\left[0; I^{-1}(\theta_0)\right]$.

Estimación máximo verosímil

Concluimos comentando cuál es la distribución asintótica del EMV. El siguiente teorema lo expresaremos supuesto que la familia de distribuciones es continua, si bien es cierto en el caso discreto con los cambios obvios.

Teorema

Sea $\{f_\theta : \theta \in \Theta\}$ una familia paramétrica de densidades con $\Theta \subseteq \mathbb{R}$. Sea una m.a.s. extraída de una población con densidad en dicha familia y supongamos que se verifican las siguientes condiciones:

- *Existen las parciales* $\dfrac{\partial^i \log\left(f_\theta(x)\right)}{\partial \theta^i}$, $i = 1, 2, 3$, $\forall \theta \in A$, *siendo* A *un abierto incluido en* Θ.
- *Existe una función* M *integrable tal que* $\left| \dfrac{\partial^3 \log\left(f_\theta(x)\right)}{\partial \theta^3} \right| < M(x)$ *con* $\mathrm{E}_\theta[M(X)] < K(\theta)$, $\forall \theta \in A$.
- *Existe al menos un valor* $\theta_0 \in A$ *verificando*
  - $\mathrm{E}_{\theta_0}\left[\dfrac{\partial \log\left(f_\theta(X)\right)}{\partial \theta}\right] = 0$, $\mathrm{E}_{\theta_0}\left[\dfrac{1}{f_\theta(X)} \dfrac{\partial^2 f_\theta(X)}{\partial \theta^2}\right] = 0$
  - $0 < I(\theta_0) < +\infty$, *con* $I(\theta)$ *la función de información de Fisher basado en una muestra de tamaño uno.*

Entonces cualquier sucesión débilmente consistente de raíces de la ecuación de verosimilitud verifica la convergencia en ley $\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{L} Z$, *donde* $Z \rightsquigarrow N_1\left[0; I^{-1}(\theta_0)\right]$.

Comentario

El resultado anterior puede generalizarse para funciones de θ. *Así, dada una función paramétrica* ψ *con derivada continua en* θ_0, *si* $\widehat{\theta}_n$ *es el EMV para* θ *se verifica que*

$$\sqrt{n}\frac{\psi(\widehat{\theta}_n) - \psi(\theta_0)}{\mid \psi'(\theta_0) \mid^{1/2}} \xrightarrow{L} Z$$

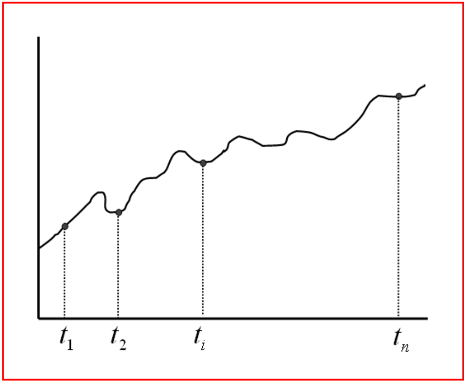
donde $Z \rightsquigarrow N_1\left[0; I^{-1}(\theta_0)\right]$.

Inferencia máximo verosímil en procesos de difusión

Sea $\{X_t; t \geq t_0\}$ un proceso de difusión, del cual conocemos sus distribuciones unidimensionales y transiciones, y para el cual es posible realizar observaciones del mismo en instantes de tiempo prefijados. Sean $t_1, \ldots, t_n$ dichos instantes de tiempo y llamemos $x_1, \ldots, x_n$ a los valores observados del proceso.

Inferencia máximo verosímil en procesos de difusión

Sea $\{X_t; t \geq t_0\}$ un proceso de difusión, del cual conocemos sus distribuciones unidimensionales y transiciones, y para el cual es posible realizar observaciones del mismo en instantes de tiempo prefijados. Sean $t_1, \ldots, t_n$ dichos instantes de tiempo y llamemos $x_1, \ldots, x_n$ a los valores observados del proceso.



Inferencia máximo verosímil en procesos de difusión

Sea $\{X_t; t \geq t_0\}$ un proceso de difusión, del cual conocemos sus distribuciones unidimensionales y transiciones, y para el cual es posible realizar observaciones del mismo en instantes de tiempo prefijados. Sean $t_1, \ldots, t_n$ dichos instantes de tiempo y llamemos $x_1, \ldots, x_n$ a los valores observados del proceso.

Sean $f_{\boldsymbol{\theta}_1}$ y $f_{\boldsymbol{\theta}}$ la densidad de la variable X_{t_1} y la función de densidad de transición (masa de probabilidad en el caso discreto), respectivamente, y sean $\boldsymbol{\theta}_1$ y $\boldsymbol{\theta}$ los parámetros asociados a ambas.

Inferencia máximo verosímil en procesos de difusión

Sea $\{X_t; t \geq t_0\}$ un proceso de difusión, del cual conocemos sus distribuciones unidimensionales y transiciones, y para el cual es posible realizar observaciones del mismo en instantes de tiempo prefijados. Sean $t_1, \ldots, t_n$ dichos instantes de tiempo y llamemos $x_1, \ldots, x_n$ a los valores observados del proceso.

Sean $f_{\boldsymbol{\theta}_1}$ y $f_{\boldsymbol{\theta}}$ la densidad de la variable X_{t_1} y la función de densidad de transición (masa de probabilidad en el caso discreto), respectivamente, y sean $\boldsymbol{\theta}_1$ y $\boldsymbol{\theta}$ los parámetros asociados a ambas.

La función de verosimilitud de la muestra observada $\mathbf{x} = (x_1, \ldots, x_n)^t$ es

$$\mathbb{L}_{\mathbf{x}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}) = f_1(x_1) \prod_{i=2}^{n} f(x_i, t_i | x_{i-1}, t_{i-1}),$$

a partir de la cual se obtendrán los estimadores máximo verosímiles de $\boldsymbol{\theta}_1$ y $\boldsymbol{\theta}$.

Inferencia máximo verosímil en procesos de difusión

Sea $\{X_t; t \geq t_0\}$ un proceso de difusión, del cual conocemos sus distribuciones unidimensionales y transiciones, y para el cual es posible realizar observaciones del mismo en instantes de tiempo prefijados. Sean $t_1, \ldots, t_n$ dichos instantes de tiempo y llamemos $x_1, \ldots, x_n$ a los valores observados del proceso.

Sean $f_{\boldsymbol{\theta}_1}$ y $f_{\boldsymbol{\theta}}$ la densidad de la variable X_{t_1} y la función de densidad de transición (masa de probabilidad en el caso discreto), respectivamente, y sean $\boldsymbol{\theta}_1$ y $\boldsymbol{\theta}$ los parámetros asociados a ambas.

La función de verosimilitud de la muestra observada $\mathbf{x} = (x_1, \ldots, x_n)^t$ es

$$\mathbb{L}_{\mathbf{x}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}) = f_1(x_1) \prod_{i=2}^{n} f(x_i, t_i | x_{i-1}, t_{i-1}),$$

a partir de la cual se obtendrán los estimadores máximo verosímiles de $\boldsymbol{\theta}_1$ y $\boldsymbol{\theta}$.

En concreto, asociadas a la muestra observada $\mathbf{x}$, se obtendrán las estimaciones $\widehat{\boldsymbol{\theta}}_1 = \widehat{\boldsymbol{\theta}}_1(\mathbf{x})$ y $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}(\mathbf{x})$ tales que

$$\mathbb{L}_{\mathbf{x}}(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}) = \underset{\boldsymbol{\theta}_1, \boldsymbol{\theta}}{Sup}\, \mathbb{L}_{\mathbf{x}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}).$$

Inferencia máximo verosímil en procesos de difusión

Sea $\{X_t; t \geq t_0\}$ un proceso de difusión, del cual conocemos sus distribuciones unidimensionales y transiciones, y para el cual es posible realizar observaciones del mismo en instantes de tiempo prefijados. Sean $t_1, \ldots, t_n$ dichos instantes de tiempo y llamemos $x_1, \ldots, x_n$ a los valores observados del proceso.

Sean $f_{\boldsymbol{\theta}_1}$ y $f_{\boldsymbol{\theta}}$ la densidad de la variable X_{t_1} y la función de densidad de transición (masa de probabilidad en el caso discreto), respectivamente, y sean $\boldsymbol{\theta}_1$ y $\boldsymbol{\theta}$ los parámetros asociados a ambas.

La función de verosimilitud de la muestra observada $\mathbf{x} = (x_1, \ldots, x_n)^t$ es

$$\mathbb{L}_{\mathbf{x}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}) = f_1(x_1) \prod_{i=2}^{n} f(x_i, t_i | x_{i-1}, t_{i-1}),$$

a partir de la cual se obtendrán los estimadores máximo verosímiles de $\boldsymbol{\theta}_1$ y $\boldsymbol{\theta}$.

En concreto, asociadas a la muestra observada $\mathbf{x}$, se obtendrán las estimaciones $\widehat{\boldsymbol{\theta}}_1 = \widehat{\boldsymbol{\theta}}_1(\mathbf{x})$ y $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}(\mathbf{x})$ tales que

$$\mathbb{L}_{\mathbf{x}}(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}) = \underset{\boldsymbol{\theta}_1, \boldsymbol{\theta}}{Sup}\, \mathbb{L}_{\mathbf{x}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}).$$

Un caso particular, pero bastante usual en la práctica, sobre todo cuando se considera una sola trayectoria del proceso, es aquél en el que la distribución de X_{t_1} es degenerada, esto es, $P(X_{t_1} = x_1) = 1$. En dicho caso la función de verosimilitud queda como

$$\mathbb{L}_{\mathbf{x}}(\boldsymbol{\theta}) = \prod_{i=2}^{n} f(x_i, t_i | x_{i-1}, t_{i-1}),$$

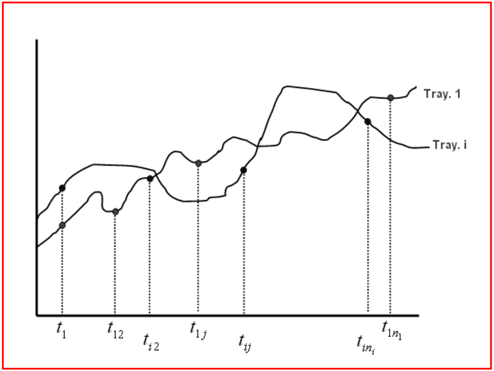
que depende sólo del parámetro $\boldsymbol{\theta}$.

Inferencia máximo verosímil en procesos de difusión

En otras ocasiones se dispone de información sobre d trayectorias, las cuales son observadas en instantes de tiempo t_{ij}, $(i = 1, \ldots, d,\ j = 1, \ldots, n_i)$. Conviene que el instante inicial sea el mismo ya que hay que imponer una distribución inicial, esto es, $t_{i1} = t_1$, $i = 1, \ldots, d$.

Inferencia máximo verosímil en procesos de difusión

En otras ocasiones se dispone de información sobre d trayectorias, las cuales son observadas en instantes de tiempo t_{ij}, $(i = 1, \ldots, d,\ j = 1, \ldots, n_i)$. Conviene que el instante inicial sea el mismo ya que hay que imponer una distribución inicial, esto es, $t_{i1} = t_1$, $i = 1, \ldots, d$.



Inferencia máximo verosímil en procesos de difusión

En otras ocasiones se dispone de información sobre d trayectorias, las cuales son observadas en instantes de tiempo t_{ij}, $(i = 1, \ldots, d,\ j = 1, \ldots, n_i)$. Conviene que el instante inicial sea el mismo ya que hay que imponer una distribución inicial, esto es, $t_{i1} = t_1$, $i = 1, \ldots, d$.

Llamando $\{x_{ij}\}_{i=1,\ldots,d;j=1,\ldots,n_i}$ a los valores observados, y $\mathbf{x}$ al vector conteniendo dichos valores, la función de verosimilitud es

- Si se considera una distribución no degenerada en t_1

$$\mathbb{L}_{\mathbf{x}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}) = \prod_{i=1}^{d} f_{\boldsymbol{\theta}_1}(x_{i1}) \prod_{j=2}^{n_i} f_{\boldsymbol{\theta}}(x_{ij}, t_{ij}|x_{i,j-1}, t_{i,j-1})$$

- Si se considera una distribución degenerada en t_1

$$\mathbb{L}_{\mathbf{x}}(\boldsymbol{\theta}) = \prod_{i=1}^{d} \prod_{j=2}^{n_i} f_{\boldsymbol{\theta}}(x_{ij}, t_{ij}|x_{i,j-1}, t_{i,j-1}).$$

Inferencia máximo verosímil en procesos de difusión

En otras ocasiones se dispone de información sobre d trayectorias, las cuales son observadas en instantes de tiempo t_{ij}, $(i = 1, \ldots, d,\ j = 1, \ldots, n_i)$. Conviene que el instante inicial sea el mismo ya que hay que imponer una distribución inicial, esto es, $t_{i1} = t_1$, $i = 1, \ldots, d$.

Llamando $\{x_{ij}\}_{i=1,\ldots,d; j=1,\ldots,n_i}$ a los valores observados, y $\mathbf{x}$ al vector conteniendo dichos valores, la función de verosimilitud es

- Si se considera una distribución no degenerada en t_1

$$\mathbb{L}_{\mathbf{x}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}) = \prod_{i=1}^{d} f_{\boldsymbol{\theta}_1}(x_{i1}) \prod_{j=2}^{n_i} f_{\boldsymbol{\theta}}(x_{ij}, t_{ij} | x_{i,j-1}, t_{i,j-1})$$

- Si se considera una distribución degenerada en t_1

$$\mathbb{L}_{\mathbf{x}}(\boldsymbol{\theta}) = \prod_{i=1}^{d} \prod_{j=2}^{n_i} f_{\boldsymbol{\theta}}(x_{ij}, t_{ij} | x_{i,j-1}, t_{i,j-1}).$$

Es habitual considerar el logaritmo de la función de verosimilitud para calcular el estimador máximo verosímil. Considerando el caso de múltiples trayectorias con distribución inicial no degenerada, ello conduce a

$$\log\left(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\theta}_1, \boldsymbol{\theta})\right) = \sum_{i=1}^{d} \log\left(f_{\boldsymbol{\theta}_1}(x_{i1})\right) + \sum_{i=1}^{d} \sum_{j=2}^{n_i} \log\left(f_{\boldsymbol{\theta}}(x_{ij}, t_{ij} | x_{i,j-1}, t_{i,j-1})\right),$$

Inferencia máximo verosímil en procesos de difusión

En otras ocasiones se dispone de información sobre d trayectorias, las cuales son observadas en instantes de tiempo t_{ij}, $(i = 1, \ldots, d,\ j = 1, \ldots, n_i)$. Conviene que el instante inicial sea el mismo ya que hay que imponer una distribución inicial, esto es, $t_{i1} = t_1$, $i = 1, \ldots, d$.

Llamando $\{x_{ij}\}_{i=1,\ldots,d;j=1,\ldots,n_i}$ a los valores observados, y $\mathbf{x}$ al vector conteniendo dichos valores, la función de verosimilitud es

- Si se considera una distribución no degenerada en t_1

$$\mathbb{L}_{\mathbf{x}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}) = \prod_{i=1}^{d} f_{\boldsymbol{\theta}_1}(x_{i1}) \prod_{j=2}^{n_i} f_{\boldsymbol{\theta}}(x_{ij}, t_{ij}|x_{i,j-1}, t_{i,j-1})$$

- Si se considera una distribución degenerada en t_1

$$\mathbb{L}_{\mathbf{x}}(\boldsymbol{\theta}) = \prod_{i=1}^{d} \prod_{j=2}^{n_i} f_{\boldsymbol{\theta}}(x_{ij}, t_{ij}|x_{i,j-1}, t_{i,j-1}).$$

Es habitual considerar el logaritmo de la función de verosimilitud para calcular el estimador máximo verosímil. Considerando el caso de múltiples trayectorias con distribución inicial no degenerada, ello conduce a

$$\log\left(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\theta}_1, \boldsymbol{\theta})\right) = \sum_{i=1}^{d} \log\left(f_{\boldsymbol{\theta}_1}(x_{i1})\right) + \sum_{i=1}^{d} \sum_{j=2}^{n_i} \log\left(f_{\boldsymbol{\theta}}(x_{ij}, t_{ij}|x_{i,j-1}, t_{i,j-1})\right),$$

por lo que, en el caso de que θ_1 y θ sean independientes, la estimación de ambos también lo será.

En tal caso, para la estimación de θ_1 sólo se considera la información del instante inicial de observación, mientras que la estimación de $\boldsymbol{\theta}$ coincide en el caso de distribución inicial degenerada y no degenerada.

Estimación máximo verosímil en el proceso Wiener

Sea $\{\widetilde{X}_t; t \geq t_0\}$ el proceso de Wiener con media infinitesimal $A_1(x,t) = \mu$ y varianza infinitesimal $A_2(x,t) = \sigma^2$, con $\mu \in \mathbb{R}$ y $\sigma > 0$.

Estimación máximo verosímil en el proceso Wiener

Sea $\{\widetilde{X}_t; t \geq t_0\}$ el proceso de Wiener con media infinitesimal $A_1(x,t) = \mu$ y varianza infinitesimal $A_2(x,t) = \sigma^2$, con $\mu \in \mathbb{R}$ y $\sigma > 0$.

Su función de densidad de transición es normal unidimensional; concretamente para $s < t$ se tiene:

$$\widetilde{X}_t | \widetilde{X}_s = y \rightsquigarrow \mathcal{N}_1 \left[y + \mu(t-s); \sigma^2(t-s) \right].$$

Estimación máximo verosímil en el proceso Wiener

Sea $\{\widetilde{X}_t; t \geq t_0\}$ el proceso de Wiener con media infinitesimal $A_1(x,t) = \mu$ y varianza infinitesimal $A_2(x,t) = \sigma^2$, con $\mu \in \mathbb{R}$ y $\sigma > 0$.

Su función de densidad de transición es normal unidimensional; concretamente para $s < t$ se tiene:

$$\widetilde{X}_t | \widetilde{X}_s = y \rightsquigarrow \mathcal{N}_1 \left[y + \mu(t-s); \sigma^2(t-s) \right].$$

Se dispone de d trayectorias, cada una con n_i datos, tomadas en los instantes de tiempo t_{ij}, $i = 1, \ldots, d$, $j = 1, \ldots, n_i$ ($t_{i1} = t_1, i = 1, \ldots, d$). Llamemos x_{ij}, $i = 1, \ldots, d$, $j = 1, \ldots, n_i$ a los valores observados y $\mathbf{x}$ al vector que contiene dichos valores. Además, supongamos que $X_{t_1} \rightsquigarrow \mathcal{N}_1[\eta, \delta]$. Sean $\boldsymbol{\gamma} = (\eta, \delta^2)^t$ y $\boldsymbol{\theta} = (\mu, \sigma^2)^t$ los vectores paramétricos correspondientes a la distribución inicial y distribuciones de transición respectivamente.

Estimación máximo verosímil en el proceso Wiener

Sea $\{\widetilde{X}_t; t \geq t_0\}$ el proceso de Wiener con media infinitesimal $A_1(x,t) = \mu$ y varianza infinitesimal $A_2(x,t) = \sigma^2$, con $\mu \in \mathbb{R}$ y $\sigma > 0$.

Su función de densidad de transición es normal unidimensional; concretamente para $s < t$ se tiene:

$$\widetilde{X}_t | \widetilde{X}_s = y \rightsquigarrow \mathcal{N}_1\left[y + \mu(t-s); \sigma^2(t-s)\right].$$

Se dispone de d trayectorias, cada una con n_i datos, tomadas en los instantes de tiempo t_{ij}, $i = 1,\ldots,d$, $j = 1,\ldots,n_i$ ($t_{i1} = t_1$,$i = 1,\ldots,d$). Llamemos x_{ij}, $i = 1,\ldots,d$, $j = 1,\ldots,n_i$ a los valores observados y $\mathbf{x}$ al vector que contiene dichos valores. Además, supongamos que $X_{t_1} \rightsquigarrow \mathcal{N}_1[\eta,\delta]$. Sean $\boldsymbol{\gamma} = (\eta,\delta^2)^t$ y $\boldsymbol{\theta} = (\mu,\sigma^2)^t$ los vectores paramétricos correspondientes a la distribución inicial y distribuciones de transición respectivamente.

Para un valor fijo de la muestra, $\mathbf{x}$, la función de verosimilitud es

$$\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}) = \prod_{i=1}^{d} f_{\boldsymbol{\gamma}}(x_{i1}) \prod_{j=2}^{n_i} f_{\boldsymbol{\theta}}(x_{ij}, t_{ij}|x_{i,j-1}, t_{i,j-1}) =$$

$$= \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi\delta^2}} \exp\left(-\frac{[x_{i1}-\eta]^2}{2\delta^2}\right) \prod_{j=2}^{n_i} \frac{1}{\sqrt{2\pi\sigma^2(t_{ij}-t_{i,j-1})}} \exp\left(-\frac{[x_{ij}-x_{i,j-1}-\mu(t_{ij}-t_{i,j-1})]^2}{2\sigma^2(t_{ij}-t_{i,j-1})}\right) =$$

$$= (2\pi\delta^2)^{-\frac{d}{2}} \exp\left(-\frac{1}{2\delta^2}\sum_{i=1}^{d}(x_{i1}-\eta)^2\right)$$

$$\times (2\pi\sigma^2)^{-\frac{-(n-d)}{2}} \prod_{i=1}^{d}\prod_{j=2}^{n_i}(t_{ij}-t_{i,j-1})^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{d}\sum_{j=2}^{n_i}(x_{i1}-x_{i,j-1}-\mu(t_{ij}-t_{i,j-1}))^2\right)$$

Estimación máximo verosímil en el proceso Wiener

El logaritmo de la función de verosimilitud es, llamando $n = n_1 + \cdots + n_d$,

$$\log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta})) = -\frac{n}{2}\log(2\pi) - \frac{d}{2}\log(\delta^2) - \frac{1}{2\delta^2}\sum_{i=1}^{d}(x_{i1} - \eta)^2$$

$$- \frac{n-d}{2}\log(\sigma^2) - \frac{1}{2}\sum_{i=1}^{d}\sum_{j=2}^{n_i}(t_{ij} - t_{i,j-1}) - \frac{1}{2\sigma^2}\sum_{i=1}^{d}\sum_{j=2}^{n_i}\frac{x_{ij} - x_{i,j-1} - \mu(t_{ij} - t_{i,j-1})}{t_{ij} - t_{i,j-1}}.$$

Estimación máximo verosímil en el proceso Wiener

El logaritmo de la función de verosimilitud es, llamando $n = n_1 + \cdots + n_d$,

$$\log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta})) = -\frac{n}{2}\log(2\pi) - \frac{d}{2}\log(\delta^2) - \frac{1}{2\delta^2}\sum_{i=1}^{d}(x_{i1} - \eta)^2$$
$$- \frac{n-d}{2}\log(\sigma^2) - \frac{1}{2}\sum_{i=1}^{d}\sum_{j=2}^{n_i}(t_{ij} - t_{i,j-1}) - \frac{1}{2\sigma^2}\sum_{i=1}^{d}\sum_{j=2}^{n_i}\frac{x_{ij} - x_{i,j-1} - \mu(t_{ij} - t_{i,j-1})}{t_{ij} - t_{i,j-1}}.$$

Supongamos que $\boldsymbol{\gamma}$ y $\boldsymbol{\theta}$ son funcionalmente independientes. Así, la estimación de ambos parámetros se puede realizar de forma independiente a partir de la expresión anterior.

Estimación máximo verosímil en el proceso Wiener

El logaritmo de la función de verosimilitud es, llamando $n = n_1 + \cdots + n_d$,

$$\log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta})) = -\frac{n}{2}\log(2\pi) - \frac{d}{2}\log(\delta^2) - \frac{1}{2\delta^2}\sum_{i=1}^{d}(x_{i1} - \eta)^2$$

$$- \frac{n-d}{2}\log(\sigma^2) - \frac{1}{2}\sum_{i=1}^{d}\sum_{j=2}^{n_i}(t_{ij} - t_{i,j-1}) - \frac{1}{2\sigma^2}\sum_{i=1}^{d}\sum_{j=2}^{n_i}\frac{x_{ij} - x_{i,j-1} - \mu(t_{ij} - t_{i,j-1})}{t_{ij} - t_{i,j-1}}.$$

Supongamos que $\boldsymbol{\gamma}$ y $\boldsymbol{\theta}$ son funcionalmente independientes. Así, la estimación de ambos parámetros se puede realizar de forma independiente a partir de la expresión anterior.

Comenzamos con $\boldsymbol{\gamma} = (\eta, \delta^2)^t$. Derivando $\log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))$ respecto de $\boldsymbol{\gamma}$ tenemos:

Estimación máximo verosímil en el proceso Wiener

El logaritmo de la función de verosimilitud es, llamando $n = n_1 + \cdots + n_d$,

$$\log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta})) = -\frac{n}{2}\log(2\pi) - \frac{d}{2}\log(\delta^2) - \frac{1}{2\delta^2}\sum_{i=1}^{d}(x_{i1}-\eta)^2$$
$$- \frac{n-d}{2}\log(\sigma^2) - \frac{1}{2}\sum_{i=1}^{d}\sum_{j=2}^{n_i}(t_{ij}-t_{i,j-1}) - \frac{1}{2\sigma^2}\sum_{i=1}^{d}\sum_{j=2}^{n_i}\frac{x_{ij}-x_{i,j-1}-\mu(t_{ij}-t_{i,j-1})}{t_{ij}-t_{i,j-1}}.$$

Supongamos que $\boldsymbol{\gamma}$ y $\boldsymbol{\theta}$ son funcionalmente independientes. Así, la estimación de ambos parámetros se puede realizar de forma independiente a partir de la expresión anterior.

Comenzamos con $\boldsymbol{\gamma} = (\eta, \delta^2)^t$. Derivando $\log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))$ respecto de $\boldsymbol{\gamma}$ tenemos:

$$\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \boldsymbol{\gamma}} = \left(\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \eta}, \frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \delta^2}\right) = \left(\frac{1}{\delta^2}\sum_{i=1}^{d}(x_{i1}-\eta), -\frac{d}{2\delta^2} + \frac{1}{2\delta^4}\sum_{i=1}^{d}(x_{i1}-\eta)^2\right)$$

Estimación máximo verosímil en el proceso Wiener

El logaritmo de la función de verosimilitud es, llamando $n = n_1 + \cdots + n_d$,

$$\log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta})) = -\frac{n}{2}\log(2\pi) - \frac{d}{2}\log(\delta^2) - \frac{1}{2\delta^2}\sum_{i=1}^{d}(x_{i1}-\eta)^2$$
$$- \frac{n-d}{2}\log(\sigma^2) - \frac{1}{2}\sum_{i=1}^{d}\sum_{j=2}^{n_i}(t_{ij}-t_{i,j-1}) - \frac{1}{2\sigma^2}\sum_{i=1}^{d}\sum_{j=2}^{n_i}\frac{x_{ij}-x_{i,j-1}-\mu(t_{ij}-t_{i,j-1})}{t_{ij}-t_{i,j-1}}.$$

Supongamos que $\boldsymbol{\gamma}$ y $\boldsymbol{\theta}$ son funcionalmente independientes. Así, la estimación de ambos parámetros se puede realizar de forma independiente a partir de la expresión anterior.

Comenzamos con $\boldsymbol{\gamma} = (\eta, \delta^2)^t$. Derivando $\log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))$ respecto de $\boldsymbol{\gamma}$ tenemos:

$$\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial \boldsymbol{\gamma}} = \left(\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial \eta}, \frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial \delta^2}\right) = \left(\frac{1}{\delta^2}\sum_{i=1}^{d}(x_{i1}-\eta), -\frac{d}{2\delta^2} + \frac{1}{2\delta^4}\sum_{i=1}^{d}(x_{i1}-\eta)^2\right)$$

El sistema de ecuaciones de verosimilitud se obtiene igualando la derivada anterior a cero, esto es

$$\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial \boldsymbol{\gamma}} = 0 \;\Leftrightarrow\; \begin{cases} \displaystyle\sum_{i=1}^{d}(x_{i1}-\eta) = 0 \\ \displaystyle\frac{1}{\delta^2}\sum_{i=1}^{d}(x_{i1}-\eta)^2 = d \end{cases}$$

Estimación máximo verosímil en el proceso Wiener

El logaritmo de la función de verosimilitud es, llamando $n = n_1 + \cdots + n_d$,

$$\log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta})) = -\frac{n}{2}\log(2\pi) - \frac{d}{2}\log(\delta^2) - \frac{1}{2\delta^2}\sum_{i=1}^{d}(x_{i1}-\eta)^2$$
$$-\frac{n-d}{2}\log(\sigma^2) - \frac{1}{2}\sum_{i=1}^{d}\sum_{j=2}^{n_i}(t_{ij}-t_{i,j-1}) - \frac{1}{2\sigma^2}\sum_{i=1}^{d}\sum_{j=2}^{n_i}\frac{x_{ij}-x_{i,j-1}-\mu(t_{ij}-t_{i,j-1})}{t_{ij}-t_{i,j-1}}.$$

Supongamos que $\boldsymbol{\gamma}$ y $\boldsymbol{\theta}$ son funcionalmente independientes. Así, la estimación de ambos parámetros se puede realizar de forma independiente a partir de la expresión anterior.

Comenzamos con $\boldsymbol{\gamma} = (\eta, \delta^2)^t$. Derivando $\log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))$ respecto de $\boldsymbol{\gamma}$ tenemos:

$$\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial \boldsymbol{\gamma}} = \left(\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial \eta}, \frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial \delta^2}\right) = \left(\frac{1}{\delta^2}\sum_{i=1}^{d}(x_{i1}-\eta), -\frac{d}{2\delta^2} + \frac{1}{2\delta^4}\sum_{i=1}^{d}(x_{i1}-\eta)^2\right)$$

El sistema de ecuaciones de verosimilitud se obtiene igualando la derivada anterior a cero, esto es

$$\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial \boldsymbol{\gamma}} = 0 \;\Leftrightarrow\; \begin{cases} \displaystyle\sum_{i=1}^{d}(x_{i1}-\eta) = 0 \\ \displaystyle\frac{1}{\delta^2}\sum_{i=1}^{d}(x_{i1}-\eta)^2 = d \end{cases}$$

La resolución del sistema proporciona

$$\widehat{\boldsymbol{\gamma}} = \left(\widehat{\eta}, \widehat{\delta}^2\right) = \left(\frac{1}{d}\sum_{i=1}^{d}x_{i1}, \frac{1}{d}\sum_{i=1}^{d}(x_{i1}-\widehat{\eta})^2\right)$$

Estimación máximo verosímil en el proceso Wiener

Comprobamos ahora que el valor obtenido que es un máximo local.Las derivadas parciales segundas son

$$\frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \eta^2} = -\frac{d}{\sigma^2},$$

Estimación máximo verosímil en el proceso Wiener

Comprobamos ahora que el valor obtenido que es un máximo local.Las derivadas parciales segundas son

$$\frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial\eta^2} = -\frac{d}{\sigma^2},\ \frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial(\delta^2)^2} = \frac{d}{2\delta^4} - \frac{1}{\delta^6}\sum_{i=1}^{n}(x_i-\eta)^2 \text{ y}$$

Estimación máximo verosímil en el proceso Wiener

Comprobamos ahora que el valor obtenido que es un máximo local.Las derivadas parciales segundas son

$$\frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial \eta^2} = -\frac{d}{\sigma^2}, \; \frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial (\delta^2)^2} = \frac{d}{2\delta^4} - \frac{1}{\delta^6}\sum_{i=1}^{n}(x_i-\eta)^2 \text{ y } \frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial \eta \partial \delta^2} = -\frac{1}{\delta^4}\sum_{i=1}^{n}(x_i-\eta)$$

Estimación máximo verosímil en el proceso Wiener

Comprobamos ahora que el valor obtenido que es un máximo local.Las derivadas parciales segundas son

$$\frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial \eta^2} = -\frac{d}{\sigma^2}, \quad \frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial (\delta^2)^2} = \frac{d}{2\delta^4} - \frac{1}{\delta^6}\sum_{i=1}^{n}(x_i-\eta)^2 \text{ y } \frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial \eta \partial \delta^2} = -\frac{1}{\delta^4}\sum_{i=1}^{n}(x_i-\eta)$$

de donde la matriz Hessiana evaluada en $\boldsymbol{\gamma} = (\eta, \delta^2)^t$ es

$$H(\boldsymbol{\gamma}) = H(\eta,\delta^2) = \frac{1}{\delta^2}\begin{pmatrix} -d & -\frac{1}{\delta^2}\sum_{i=1}^{d}(x_{i1}-\eta) \\ -\frac{1}{\delta^2}\sum_{i=1}^{d}(x_{i1}-\eta) & -\frac{d}{2\delta^2} - \frac{1}{\delta^4}\sum_{i=1}^{d}(x_{i1}-\eta)^2 \end{pmatrix}$$

Estimación máximo verosímil en el proceso Wiener

Comprobamos ahora que el valor obtenido que es un máximo local.Las derivadas parciales segundas son

$$\frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial\eta^2} = -\frac{d}{\sigma^2},\ \frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial(\delta^2)^2} = \frac{d}{2\delta^4} - \frac{1}{\delta^6}\sum_{i=1}^{n}(x_i-\eta)^2 \text{ y } \frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial\eta\partial\delta^2} = -\frac{1}{\delta^4}\sum_{i=1}^{n}(x_i-\eta)$$

de donde la matriz Hessiana evaluada en $\boldsymbol{\gamma} = (\eta, \delta^2)^t$ es

$$H(\boldsymbol{\gamma}) = H(\eta,\delta^2) = \frac{1}{\delta^2}\begin{pmatrix} -d & -\frac{1}{\delta^2}\sum_{i=1}^{d}(x_{i1}-\eta) \\ -\frac{1}{\delta^2}\sum_{i=1}^{d}(x_{i1}-\eta) & -\frac{d}{2\delta^2} - \frac{1}{\delta^4}\sum_{i=1}^{d}(x_{i1}-\eta)^2 \end{pmatrix}$$

y en el punto $\widehat{\boldsymbol{\gamma}} = (\widehat{\eta}, \widehat{\delta}^2)^t$ es

$$H(\widehat{\boldsymbol{\gamma}}) = H(\widehat{\eta},\widehat{\delta}^2) = -\frac{d}{\widehat{\delta}^2}\begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2\widehat{\delta}^2} \end{pmatrix}$$

que se puede comprobar es definida negativa.

Estimación máximo verosímil en el proceso Wiener

Comprobamos ahora que el valor obtenido que es un máximo local.Las derivadas parciales segundas son

$$\frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial\eta^2} = -\frac{d}{\sigma^2},\ \frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial(\delta^2)^2} = \frac{d}{2\delta^4} - \frac{1}{\delta^6}\sum_{i=1}^{n}(x_i-\eta)^2 \text{ y } \frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial\eta\partial\delta^2} = -\frac{1}{\delta^4}\sum_{i=1}^{n}(x_i-\eta)$$

de donde la matriz Hessiana evaluada en $\boldsymbol{\gamma} = (\eta,\delta^2)^t$ es

$$H(\boldsymbol{\gamma}) = H(\eta,\delta^2) = \frac{1}{\delta^2}\begin{pmatrix} -d & -\dfrac{1}{\delta^2}\displaystyle\sum_{i=1}^{d}(x_{i1}-\eta) \\ -\dfrac{1}{\delta^2}\displaystyle\sum_{i=1}^{d}(x_{i1}-\eta) & -\dfrac{d}{2\delta^2} - \dfrac{1}{\delta^4}\displaystyle\sum_{i=1}^{d}(x_{i1}-\eta)^2 \end{pmatrix}$$

y en el punto $\widehat{\boldsymbol{\gamma}} = (\widehat{\eta},\widehat{\delta}^2)^t$ es

$$H(\widehat{\boldsymbol{\gamma}}) = H(\widehat{\eta},\widehat{\delta}^2) = -\frac{d}{\widehat{\delta}^2}\begin{pmatrix} 1 & 0 \\ 0 & \dfrac{1}{2\widehat{\delta}^2} \end{pmatrix}$$

que se puede comprobar es definida negativa.

No obstante, esto demuestra que es un máximo local, pero no global. Ello habría que hacerlo tomando límites de la función de verosimilitud en la frontera de $\Theta = \mathbb{R}\times\mathbb{R}^+$, pudiéndose comprobar que esos límites son cero.

Estimación máximo verosímil en el proceso Wiener

Procedamos ahora con con $\boldsymbol{\theta} = (\mu, \sigma^2)^t$. Derivando $\log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))$ respecto de $\boldsymbol{\theta}$ tenemos:

Estimación máximo verosímil en el proceso Wiener

Procedamos ahora con con $\boldsymbol{\theta} = (\mu, \sigma^2)^t$. Derivando $\log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))$ respecto de $\boldsymbol{\theta}$ tenemos:

$$\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} =$$

Estimación máximo verosímil en el proceso Wiener

Procedamos ahora con con $\boldsymbol{\theta} = (\mu, \sigma^2)^t$. Derivando $\log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))$ respecto de $\boldsymbol{\theta}$ tenemos:

$$\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} = \left(\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \mu}, \right.$$

Estimación máximo verosímil en el proceso Wiener

Procedamos ahora con con $\boldsymbol{\theta} = (\mu, \sigma^2)^t$. Derivando $\log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))$ respecto de $\boldsymbol{\theta}$ tenemos:

$$\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} = \left(\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \mu}, \frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \sigma^2} \right)$$

Estimación máximo verosímil en el proceso Wiener

Procedamos ahora con con $\boldsymbol{\theta} = (\mu, \sigma^2)^t$. Derivando $\log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))$ respecto de $\boldsymbol{\theta}$ tenemos:

$$\begin{aligned}
\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} &= \left(\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \mu}, \frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \sigma^2} \right) \\
&= \left(\frac{1}{\sigma^2} \sum_{i=1}^{d} \sum_{j=2}^{n_i} (x_{i1} - x_{i,j-1} - \mu(t_{ij} - t_{i,j-1})), \right.
\end{aligned}$$

Estimación máximo verosímil en el proceso Wiener

Procedamos ahora con con $\boldsymbol{\theta} = (\mu, \sigma^2)^t$. Derivando $\log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))$ respecto de $\boldsymbol{\theta}$ tenemos:

$$
\begin{aligned}
\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} &= \left(\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \mu}, \frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \sigma^2} \right) \\
&= \left(\frac{1}{\sigma^2} \sum_{i=1}^{d} \sum_{j=2}^{n_i} (x_{i1} - x_{i,j-1} - \mu(t_{ij} - t_{i,j-1})), \; -\frac{n-d}{2\delta^2} + \frac{1}{2\delta^4} \sum_{i=1}^{d} \sum_{j=2}^{n_i} (x_{i-} - x_{i,j-1} - \mu(t_{ij} - t_{i,j-1}))^2 \right)
\end{aligned}
$$

Estimación máximo verosímil en el proceso Wiener

Procedamos ahora con con $\boldsymbol{\theta} = (\mu, \sigma^2)^t$. Derivando $\log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))$ respecto de $\boldsymbol{\theta}$ tenemos:

$$\begin{aligned}\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} &= \left(\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \mu}, \frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \sigma^2}\right) \\ &= \left(\frac{1}{\sigma^2}\sum_{i=1}^{d}\sum_{j=2}^{n_i}(x_{i1} - x_{i,j-1} - \mu(t_{ij} - t_{i,j-1})), \; -\frac{n-d}{2\delta^2} + \frac{1}{2\delta^4}\sum_{i=1}^{d}\sum_{j=2}^{n_i}(x_{i-} - x_{i,j-1} - \mu(t_{ij} - t_{i,j-1}))^2\right) \\ &= \left(\frac{1}{\sigma^2}\sum_{i=1}^{d}(x_{i,n_i} - x_{i1} - \mu(t_{i,n_i} - t_{i1})),\right.\end{aligned}$$

Estimación máximo verosímil en el proceso Wiener

Procedamos ahora con con $\boldsymbol{\theta} = (\mu, \sigma^2)^t$. Derivando $\log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))$ respecto de $\boldsymbol{\theta}$ tenemos:

$$\begin{aligned}
\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} &= \left(\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \mu}, \frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \sigma^2} \right) \\
&= \left(\frac{1}{\sigma^2} \sum_{i=1}^{d} \sum_{j=2}^{n_i} (x_{i1} - x_{i,j-1} - \mu(t_{ij} - t_{i,j-1})), \; -\frac{n-d}{2\delta^2} + \frac{1}{2\delta^4} \sum_{i=1}^{d} \sum_{j=2}^{n_i} (x_{i-} - x_{i,j-1} - \mu(t_{ij} - t_{i,j-1}))^2 \right) \\
&= \left(\frac{1}{\sigma^2} \sum_{i=1}^{d} (x_{i,n_i} - x_{i1} - \mu(t_{i,n_i} - t_{i1})), \; -\frac{n-d}{2\delta^2} + \frac{1}{2\delta^4} \sum_{i=1}^{d} \sum_{j=2}^{n_i} \frac{(x_{i-} - x_{i,j-1} - \mu(t_{ij} - t_{i,j-1}))^2}{t_{ij} - t_{i,j-1}} \right).
\end{aligned}$$

Estimación máximo verosímil en el proceso Wiener

Procedamos ahora con con $\boldsymbol{\theta} = (\mu, \sigma^2)^t$. Derivando $\log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))$ respecto de $\boldsymbol{\theta}$ tenemos:

$$\begin{aligned}\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} &= \left(\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \mu}, \frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \sigma^2}\right) \\ &= \left(\frac{1}{\sigma^2}\sum_{i=1}^{d}\sum_{j=2}^{n_i}(x_{i1} - x_{i,j-1} - \mu(t_{ij} - t_{i,j-1})), \; -\frac{n-d}{2\delta^2} + \frac{1}{2\delta^4}\sum_{i=1}^{d}\sum_{j=2}^{n_i}(x_{i-} - x_{i,j-1} - \mu(t_{ij} - t_{i,j-1}))^2\right) \\ &= \left(\frac{1}{\sigma^2}\sum_{i=1}^{d}(x_{i,n_i} - x_{i1} - \mu(t_{i,n_i} - t_{i1})), \; -\frac{n-d}{2\delta^2} + \frac{1}{2\delta^4}\sum_{i=1}^{d}\sum_{j=2}^{n_i}\frac{(x_{i-} - x_{i,j-1} - \mu(t_{ij} - t_{i,j-1}))^2}{t_{ij} - t_{i,j-1}}\right).\end{aligned}$$

El sistema de ecuaciones de verosimilitud se obtiene igualando la derivada anterior a cero, esto es

$$\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} = 0 \;\Leftrightarrow\; \begin{cases} \displaystyle\sum_{i=1}^{d}(x_{i,n_i} - x_{i1}) = \mu\sum_{i=1}^{d}(t_{i,n_i} - t_{i1}) \\ \displaystyle\frac{1}{\sigma^2}\sum_{i=1}^{d}\sum_{j=2}^{n_i}\frac{(x_{ij} - x_{i,j-1} - \mu(t_{ij} - t_{i,j-1}))^2}{t_{ij} - t_{i,j-1}} = n - d \end{cases}$$

Estimación máximo verosímil en el proceso Wiener

Procedamos ahora con con $\boldsymbol{\theta} = (\mu, \sigma^2)^t$. Derivando $\log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))$ respecto de $\boldsymbol{\theta}$ tenemos:

$$\begin{aligned}
\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} &= \left(\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \mu}, \frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \sigma^2} \right) \\
&= \left(\frac{1}{\sigma^2} \sum_{i=1}^{d} \sum_{j=2}^{n_i} (x_{i1} - x_{i,j-1} - \mu(t_{ij} - t_{i,j-1})), \; -\frac{n-d}{2\delta^2} + \frac{1}{2\delta^4} \sum_{i=1}^{d} \sum_{j=2}^{n_i} (x_{i-} - x_{i,j-1} - \mu(t_{ij} - t_{i,j-1}))^2 \right) \\
&= \left(\frac{1}{\sigma^2} \sum_{i=1}^{d} (x_{i,n_i} - x_{i1} - \mu(t_{i,n_i} - t_{i1})), \; -\frac{n-d}{2\delta^2} + \frac{1}{2\delta^4} \sum_{i=1}^{d} \sum_{j=2}^{n_i} \frac{(x_{i-} - x_{i,j-1} - \mu(t_{ij} - t_{i,j-1}))^2}{t_{ij} - t_{i,j-1}} \right).
\end{aligned}$$

El sistema de ecuaciones de verosimilitud se obtiene igualando la derivada anterior a cero, esto es

$$\frac{\partial \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma}, \boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} = 0 \;\Leftrightarrow\; \left\{ \begin{array}{l} \displaystyle\sum_{i=1}^{d} (x_{i,n_i} - x_{i1}) = \mu \sum_{i=1}^{d} (t_{i,n_i} - t_{i1}) \\ \displaystyle\frac{1}{\sigma^2} \sum_{i=1}^{d} \sum_{j=2}^{n_i} \frac{(x_{ij} - x_{i,j-1} - \mu(t_{ij} - t_{i,j-1}))^2}{t_{ij} - t_{i,j-1}} = n - d \end{array} \right.$$

La resolución del sistema proporciona

$$\widehat{\gamma} = \left(\widehat{\mu}, \widehat{\sigma^2} \right) = \left(\frac{\displaystyle\sum_{i=1}^{d} (x_{i,n_i} - x_{i1})}{\displaystyle\sum_{i=1}^{d} (t_{i,n_i} - t_{i1})}, \frac{1}{n-d} \sum_{i=1}^{d} \sum_{j=2}^{n_i} \frac{(x_{ij} - x_{i,j-1} - \widehat{\mu}(t_{ij} - t_{i,j-1}))^2}{t_{ij} - t_{i,j-1}} \right)$$

Estimación máximo verosímil en el proceso Wiener

Comprobamos ahora que el valor obtenido que es un máximo local.Las derivadas parciales segundas son

$$\frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial\mu^2} = -\frac{1}{\sigma^2}\sum_{i=1}^{d}(t_{i,n_i} - t_{i1}),$$

Estimación máximo verosímil en el proceso Wiener

Comprobamos ahora que el valor obtenido que es un máximo local.Las derivadas parciales segundas son

$$\frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial\mu^2} = -\frac{1}{\sigma^2}\sum_{i=1}^{d}(t_{i,n_i}-t_{i1}),\ \frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial(\sigma^2)^2} = \frac{n-d}{2\delta^4} - \frac{1}{\delta^6}\sum_{i=1}^{d}\sum_{j=2}^{n_i}\frac{(x_{i-}-x_{i,j-1}-\mu(t_{ij}-t_{i,j-1}))^2}{t_{ij}-t_{i,j-1}}$$ y

$$\frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial\eta\partial\delta^2} = -\frac{1}{\delta^4}\sum_{i=1}^{d}\sum_{j=2}^{n_i}(x_{ij}-x_{i,j-1}-\mu(t_{ij}-t_{i,j-1})$$

Estimación máximo verosímil en el proceso Wiener

Comprobamos ahora que el valor obtenido que es un máximo local.Las derivadas parciales segundas son

$$\frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial\mu^2} = -\frac{1}{\sigma^2}\sum_{i=1}^{d}(t_{i,n_i}-t_{i1}),\ \frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial(\sigma^2)^2} = \frac{n-d}{2\delta^4} - \frac{1}{\delta^6}\sum_{i=1}^{d}\sum_{j=2}^{n_i}\frac{(x_{i-}-x_{i,j-1}-\mu(t_{ij}-t_{i,j-1}))^2}{t_{ij}-t_{i,j-1}}$$ y

$$\frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial\eta\partial\delta^2} = -\frac{1}{\delta^4}\sum_{i=1}^{d}\sum_{j=2}^{n_i}(x_{ij}-x_{i,j-1}-\mu(t_{ij}-t_{i,j-1})$$

de donde la matriz Hessiana evaluada en $\boldsymbol{\theta} = (\mu,\sigma^2)^t$ es

$$H(\boldsymbol{\theta}) = H(\mu,\sigma^2) = \frac{1}{\sigma^2}\begin{pmatrix} -\sum_{i=1}^{d}(t_{i,n_i}-t_{i1}) & -\frac{1}{\delta^2}\sum_{i=1}^{d}\sum_{j=2}^{n_i}(x_{ij}-x_{i,j-1}-\mu(t_{ij}-t_{i,j-1}) \\ -\frac{1}{\delta^2}\sum_{i=1}^{d}\sum_{j=2}^{n_i}(x_{ij}-x_{i,j-1}-\mu(t_{ij}-t_{i,j-1}) & -\frac{n-d}{2\delta^2} - \frac{1}{\delta^4}\sum_{j=2}^{n_i}\frac{(x_{i-}-x_{i,j-1}-\mu(t_{ij}-t_{i,j-1}))^2}{t_{ij}-t_{i,j-1}} \end{pmatrix}$$

Estimación máximo verosímil en el proceso Wiener

Comprobamos ahora que el valor obtenido que es un máximo local. Las derivadas parciales segundas son

$$\frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial \mu^2} = -\frac{1}{\sigma^2}\sum_{i=1}^{d}(t_{i,n_i}-t_{i1}),\ \frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial (\sigma^2)^2} = \frac{n-d}{2\delta^4} - \frac{1}{\delta^6}\sum_{i=1}^{d}\sum_{j=2}^{n_i}\frac{(x_{i-}-x_{i,j-1}-\mu(t_{ij}-t_{i,j-1}))^2}{t_{ij}-t_{i,j-1}} \text{ y}$$

$$\frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial \eta \partial \delta^2} = -\frac{1}{\delta^4}\sum_{i=1}^{d}\sum_{j=2}^{n_i}(x_{ij}-x_{i,j-1}-\mu(t_{ij}-t_{i,j-1})$$

de donde la matriz Hessiana evaluada en $\boldsymbol{\theta} = (\mu,\sigma^2)^t$ es

$$H(\boldsymbol{\theta}) = H(\mu,\sigma^2) = \frac{1}{\sigma^2}\begin{pmatrix} -\sum_{i=1}^{d}(t_{i,n_i}-t_{i1}) & -\frac{1}{\delta^2}\sum_{i=1}^{d}\sum_{j=2}^{n_i}(x_{ij}-x_{i,j-1}-\mu(t_{ij}-t_{i,j-1}) \\ -\frac{1}{\delta^2}\sum_{i=1}^{d}\sum_{j=2}^{n_i}(x_{ij}-x_{i,j-1}-\mu(t_{ij}-t_{i,j-1}) & -\frac{n-d}{2\delta^2} - \frac{1}{\delta^4}\sum_{j=2}^{n_i}\frac{(x_{i-}-x_{i,j-1}-\mu(t_{ij}-t_{i,j-1}))^2}{t_{ij}-t_{i,j-1}} \end{pmatrix}$$

y en el punto $\widehat{\boldsymbol{\theta}} = (\widehat{\mu},\widehat{\sigma}^2)^t$ es

$$H(\widehat{\boldsymbol{\theta}}) = H(\widehat{\mu},\widehat{\sigma}^2) = -\frac{1}{\widehat{\sigma}^2}\begin{pmatrix} \sum_{i=1}^{d}(t_{i,n_i}-t_{i1}) & 0 \\ 0 & \frac{1}{2\widehat{\sigma}^2} \end{pmatrix}$$

que se puede comprobar es definida negativa.

Estimación máximo verosímil en el proceso Wiener

Comprobamos ahora que el valor obtenido que es un máximo local.Las derivadas parciales segundas son

$$\frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial \mu^2} = -\frac{1}{\sigma^2}\sum_{i=1}^{d}(t_{i,n_i}-t_{i1}),\quad \frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial (\sigma^2)^2} = \frac{n-d}{2\delta^4} - \frac{1}{\delta^6}\sum_{i=1}^{d}\sum_{j=2}^{n_i}\frac{(x_{i-}-x_{i,j-1}-\mu(t_{ij}-t_{i,j-1}))^2}{t_{ij}-t_{i,j-1}} \text{ y}$$

$$\frac{\partial^2 \log(\mathbb{L}_{\mathbf{x}}(\boldsymbol{\gamma},\boldsymbol{\theta}))}{\partial \eta \partial \delta^2} = -\frac{1}{\delta^4}\sum_{i=1}^{d}\sum_{j=2}^{n_i}(x_{ij}-x_{i,j-1}-\mu(t_{ij}-t_{i,j-1})$$

de donde la matriz Hessiana evaluada en $\boldsymbol{\theta} = (\mu,\sigma^2)^t$ es

$$H(\boldsymbol{\theta}) = H(\mu,\sigma^2) = \frac{1}{\sigma^2}\begin{pmatrix} -\sum_{i=1}^{d}(t_{i,n_i}-t_{i1}) & -\frac{1}{\delta^2}\sum_{i=1}^{d}\sum_{j=2}^{n_i}(x_{ij}-x_{i,j-1}-\mu(t_{ij}-t_{i,j-1}) \\ -\frac{1}{\delta^2}\sum_{i=1}^{d}\sum_{j=2}^{n_i}(x_{ij}-x_{i,j-1}-\mu(t_{ij}-t_{i,j-1}) & -\frac{n-d}{2\delta^2} - \frac{1}{\delta^4}\sum_{j=2}^{n_i}\frac{(x_{i-}-x_{i,j-1}-\mu(t_{ij}-t_{i,j-1}))^2}{t_{ij}-t_{i,j-1}} \end{pmatrix}$$

y en el punto $\widehat{\boldsymbol{\theta}} = (\widehat{\mu},\widehat{\sigma}^2)^t$ es

$$H(\widehat{\boldsymbol{\theta}}) = H(\widehat{\mu},\widehat{\sigma}^2) = -\frac{1}{\widehat{\sigma}^2}\begin{pmatrix} \sum_{i=1}^{d}(t_{i,n_i}-t_{i1}) & 0 \\ 0 & \frac{1}{2\widehat{\sigma}^2} \end{pmatrix}$$

que se puede comprobar es definida negativa.

No obstante, esto demuestra que es un máximo local, pero no global. Ello habría que hacerlo tomando límites de la función de verosimilitud en la frontera de $\Theta = \mathbb{R}\times\mathbb{R}^+$, pudiéndose comprobar que esos límites son cero.

Estimación máximo verosímil en el proceso Wiener

A partir de la estimación máximo verosímil de los parámetros es posible, mediante la aplicación del teorema de Zehna, estimar funciones paramétricas asociadas al proceso como, por ejemplo, la tendencia (función media), la varianza o la covarianza. Así, aplicando las propiedades del condicionamiento se tiene

Estimación máximo verosímil en el proceso Wiener

A partir de la estimación máximo verosímil de los parámetros es posible, mediante la aplicación del teorema de Zehna, estimar funciones paramétricas asociadas al proceso como, por ejemplo, la tendencia (función media), la varianza o la covarianza. Así, aplicando las propiedades del condicionamiento se tiene

$$E\left[X_t^k\right] = E\left[E\left[X_t^k | X_{t_1}\right]\right], \; t \geq t_1; \; k \geq 1$$

Estimación máximo verosímil en el proceso Wiener

A partir de la estimación máximo verosímil de los parámetros es posible, mediante la aplicación del teorema de Zehna, estimar funciones paramétricas asociadas al proceso como, por ejemplo, la tendencia (función media), la varianza o la covarianza. Así, aplicando las propiedades del condicionamiento se tiene

$$E\left[X_t^k\right] = E\left[E\left[X_t^k | X_{t_1}\right]\right], \, t \geq t_1; \, k \geq 1$$

y como $E\left[X_t^k | X_{t_1}\right]$ es un variable aleatoria cuyos valores son $E\left[X_t^k | X_{t_1} = y\right]$, entonces, teniendo en cuenta la expresión de las densidades de transición, se concluye

Estimación máximo verosímil en el proceso Wiener

A partir de la estimación máximo verosímil de los parámetros es posible, mediante la aplicación del teorema de Zehna, estimar funciones paramétricas asociadas al proceso como, por ejemplo, la tendencia (función media), la varianza o la covarianza. Así, aplicando las propiedades del condicionamiento se tiene

$$E\left[X_t^k\right] = E\left[E\left[X_t^k | X_{t_1}\right]\right], \, t \geq t_1; \, k \geq 1$$

y como $E\left[X_t^k | X_{t_1}\right]$ es un variable aleatoria cuyos valores son $E\left[X_t^k | X_{t_1} = y\right]$, entonces, teniendo en cuenta la expresión de las densidades de transición, se concluye

$$E\left[X_t\right] = E\left[E\left[X_t | X_{t_1}\right]\right] = E\left[X_{t_1} + \mu(t - t_1)\right] = E\left[X_{t_1}\right] + \mu(t - t_1)$$

Estimación máximo verosímil en el proceso Wiener

A partir de la estimación máximo verosímil de los parámetros es posible, mediante la aplicación del teorema de Zehna, estimar funciones paramétricas asociadas al proceso como, por ejemplo, la tendencia (función media), la varianza o la covarianza. Así, aplicando las propiedades del condicionamiento se tiene

$$E\left[X_t^k\right] = E\left[E\left[X_t^k | X_{t_1}\right]\right],\ t \geq t_1;\ k \geq 1$$

y como $E\left[X_t^k | X_{t_1}\right]$ es un variable aleatoria cuyos valores son $E\left[X_t^k | X_{t_1} = y\right]$, entonces, teniendo en cuenta la expresión de las densidades de transición, se concluye

$$E\left[X_t\right] = E\left[E\left[X_t | X_{t_1}\right]\right] = E\left[X_{t_1} + \mu(t - t_1)\right] = E\left[X_{t_1}\right] + \mu(t - t_1)$$

$$E\left[X_t^2\right] = E\left[E\left[X_t^2 | X_{t_1}\right]\right] = E\left[\sigma^2(t - t_1) + \left(X_{t_1} + \mu(t - t_1)\right)^2\right] =$$

Estimación máximo verosímil en el proceso Wiener

A partir de la estimación máximo verosímil de los parámetros es posible, mediante la aplicación del teorema de Zehna, estimar funciones paramétricas asociadas al proceso como, por ejemplo, la tendencia (función media), la varianza o la covarianza. Así, aplicando las propiedades del condicionamiento se tiene

$$E\left[X_t^k\right] = E\left[E\left[X_t^k | X_{t_1}\right]\right],\ t \geq t_1;\ k \geq 1$$

y como $E\left[X_t^k | X_{t_1}\right]$ es un variable aleatoria cuyos valores son $E\left[X_t^k | X_{t_1} = y\right]$, entonces, teniendo en cuenta la expresión de las densidades de transición, se concluye

$$E\left[X_t\right] = E\left[E\left[X_t | X_{t_1}\right]\right] = E\left[X_{t_1} + \mu(t-t_1)\right] = E\left[X_{t_1}\right] + \mu(t-t_1)$$

$$E\left[X_t^2\right] = E\left[E\left[X_t^2 | X_{t_1}\right]\right] = E\left[\sigma^2(t-t_1) + (X_{t_1} + \mu(t-t_1))^2\right] = \sigma^2(t-t_1) + E[X_{t_1}^2] + \mu^2(t-t_1)^2 + 2E[X_{t_1}]\mu(t-t_1)$$

Estimación máximo verosímil en el proceso Wiener

A partir de la estimación máximo verosímil de los parámetros es posible, mediante la aplicación del teorema de Zehna, estimar funciones paramétricas asociadas al proceso como, por ejemplo, la tendencia (función media), la varianza o la covarianza. Así, aplicando las propiedades del condicionamiento se tiene

$$E\left[X_t^k\right] = E\left[E\left[X_t^k | X_{t_1}\right]\right],\ t \geq t_1;\ k \geq 1$$

y como $E\left[X_t^k | X_{t_1}\right]$ es un variable aleatoria cuyos valores son $E\left[X_t^k | X_{t_1} = y\right]$, entonces, teniendo en cuenta la expresión de las densidades de transición, se concluye

$$E\left[X_t\right] = E\left[E\left[X_t | X_{t_1}\right]\right] = E\left[X_{t_1} + \mu(t-t_1)\right] = E\left[X_{t_1}\right] + \mu(t-t_1)$$

$$E\left[X_t^2\right] = E\left[E\left[X_t^2 | X_{t_1}\right]\right] = E\left[\sigma^2(t-t_1) + (X_{t_1} + \mu(t-t_1))^2\right] = \sigma^2(t-t_1) + E[X_{t_1}^2] + \mu^2(t-t_1)^2 + 2E[X_{t_1}]\mu(t-t_1)$$

$$\mathrm{Var}[X_t] = E\left[X_t^2\right] - \left(E\left[X_t\right]\right)^2 = \sigma^2(t-t_1) + E[X_{t_1}^2] - \left(E\left[X_{t_1}\right]\right)^2 = \sigma^2(t-t_1) + \mathrm{Var}[X(t_1)].$$

Estimación máximo verosímil en el proceso Wiener

A partir de la estimación máximo verosímil de los parámetros es posible, mediante la aplicación del teorema de Zehna, estimar funciones paramétricas asociadas al proceso como, por ejemplo, la tendencia (función media), la varianza o la covarianza. Así, aplicando las propiedades del condicionamiento se tiene

$$E\left[X_t^k\right] = E\left[E\left[X_t^k | X_{t_1}\right]\right], \; t \geq t_1; \; k \geq 1$$

y como $E\left[X_t^k | X_{t_1}\right]$ es un variable aleatoria cuyos valores son $E\left[X_t^k | X_{t_1} = y\right]$, entonces, teniendo en cuenta la expresión de las densidades de transición, se concluye

$$E\left[X_t\right] = E\left[E\left[X_t | X_{t_1}\right]\right] = E\left[X_{t_1} + \mu(t-t_1)\right] = E\left[X_{t_1}\right] + \mu(t-t_1)$$

$$E\left[X_t^2\right] = E\left[E\left[X_t^2 | X_{t_1}\right]\right] = E\left[\sigma^2(t-t_1) + (X_{t_1} + \mu(t-t_1))^2\right] = \sigma^2(t-t_1) + E[X_{t_1}^2] + \mu^2(t-t_1)^2 + 2E[X_{t_1}]\mu(t-t_1)$$

$$\mathrm{Var}[X_t] = E\left[X_t^2\right] - \left(E\left[X_t\right]\right)^2 = \sigma^2(t-t_1) + E[X_{t_1}^2] - \left(E\left[X_{t_1}\right]\right)^2 = \sigma^2(t-t_1) + \mathrm{Var}[X(t_1)].$$

Observemos que la tendencia y la varianza, a partir del primer instante de observación, dependen de los momentos de X_{t_1}. En este caso concreto, y dado que hemos supuesto una distribución normal en t_1, con media η y varianza δ^2, entonces

$$E\left[X_t\right] = \eta + \mu(t-t_1)$$

Estimación máximo verosímil en el proceso Wiener

A partir de la estimación máximo verosímil de los parámetros es posible, mediante la aplicación del teorema de Zehna, estimar funciones paramétricas asociadas al proceso como, por ejemplo, la tendencia (función media), la varianza o la covarianza. Así, aplicando las propiedades del condicionamiento se tiene

$$E\left[X_t^k\right] = E\left[E\left[X_t^k|X_{t_1}\right]\right], \; t \geq t_1; \; k \geq 1$$

y como $E\left[X_t^k|X_{t_1}\right]$ es un variable aleatoria cuyos valores son $E\left[X_t^k|X_{t_1}=y\right]$, entonces, teniendo en cuenta la expresión de las densidades de transición, se concluye

$$E\left[X_t\right] = E\left[E\left[X_t|X_{t_1}\right]\right] = E\left[X_{t_1}+\mu(t-t_1)\right] = E\left[X_{t_1}\right]+\mu(t-t_1)$$

$$E\left[X_t^2\right] = E\left[E\left[X_t^2|X_{t_1}\right]\right] = E\left[\sigma^2(t-t_1)+\left(X_{t_1}+\mu(t-t_1)\right)^2\right] = \sigma^2(t-t_1)+E[X_{t_1}^2]+\mu^2(t-t_1)^2+2E[X_{t_1}]\mu(t-t_1)$$

$$\mathrm{Var}[X_t] = E\left[X_t^2\right]-\left(E\left[X_t\right]\right)^2 = \sigma^2(t-t_1)+E[X_{t_1}^2]-\left(E\left[X_{t_1}\right]\right)^2 = \sigma^2(t-t_1)+\mathrm{Var}[X(t_1)].$$

Observemos que la tendencia y la varianza, a partir del primer instante de observación, dependen de los momentos de X_{t_1}. En este caso concreto, y dado que hemos supuesto una distribución normal en t_1, con media η y varianza δ^2, entonces

$$E\left[X_t\right] = \eta+\mu(t-t_1)$$

$$\mathrm{Var}[X_t] = \delta^2+\sigma^2(t-t_1)$$

por lo que, a partir de una muestra observada, sus estimaciones máximo verosímiles son

Estimación máximo verosímil en el proceso Wiener

A partir de la estimación máximo verosímil de los parámetros es posible, mediante la aplicación del teorema de Zehna, estimar funciones paramétricas asociadas al proceso como, por ejemplo, la tendencia (función media), la varianza o la covarianza. Así, aplicando las propiedades del condicionamiento se tiene

$$E\left[X_t^k\right] = E\left[E\left[X_t^k | X_{t_1}\right]\right], \; t \geq t_1; \; k \geq 1$$

y como $E\left[X_t^k | X_{t_1}\right]$ es un variable aleatoria cuyos valores son $E\left[X_t^k | X_{t_1} = y\right]$, entonces, teniendo en cuenta la expresión de las densidades de transición, se concluye

$$E\left[X_t\right] = E\left[E\left[X_t | X_{t_1}\right]\right] = E\left[X_{t_1} + \mu(t - t_1)\right] = E\left[X_{t_1}\right] + \mu(t - t_1)$$

$$E\left[X_t^2\right] = E\left[E\left[X_t^2 | X_{t_1}\right]\right] = E\left[\sigma^2(t - t_1) + (X_{t_1} + \mu(t - t_1))^2\right] = \sigma^2(t - t_1) + E[X_{t_1}^2] + \mu^2(t - t_1)^2 + 2E[X_{t_1}]\mu(t - t_1)$$

$$\mathrm{Var}[X_t] = E\left[X_t^2\right] - \left(E\left[X_t\right]\right)^2 = \sigma^2(t - t_1) + E[X_{t_1}^2] - \left(E\left[X_{t_1}\right]\right)^2 = \sigma^2(t - t_1) + \mathrm{Var}[X(t_1)].$$

Observemos que la tendencia y la varianza, a partir del primer instante de observación, dependen de los momentos de X_{t_1}. En este caso concreto, y dado que hemos supuesto una distribución normal en t_1, con media η y varianza δ^2, entonces

$$E\left[X_t\right] = \eta + \mu(t - t_1)$$

$$\mathrm{Var}[X_t] = \delta^2 + \sigma^2(t - t_1)$$

por lo que, a partir de una muestra observada, sus estimaciones máximo verosímiles son

$$\widehat{E\left[X_t\right]} = \widehat{\eta} + \widehat{\mu}(t - t_1)$$

Estimación máximo verosímil en el proceso Wiener

A partir de la estimación máximo verosímil de los parámetros es posible, mediante la aplicación del teorema de Zehna, estimar funciones paramétricas asociadas al proceso como, por ejemplo, la tendencia (función media), la varianza o la covarianza. Así, aplicando las propiedades del condicionamiento se tiene

$$E\left[X_t^k\right] = E\left[E\left[X_t^k | X_{t_1}\right]\right],\ t \geq t_1;\ k \geq 1$$

y como $E\left[X_t^k | X_{t_1}\right]$ es un variable aleatoria cuyos valores son $E\left[X_t^k | X_{t_1} = y\right]$, entonces, teniendo en cuenta la expresión de las densidades de transición, se concluye

$$E\left[X_t\right] = E\left[E\left[X_t | X_{t_1}\right]\right] = E\left[X_{t_1} + \mu(t - t_1)\right] = E\left[X_{t_1}\right] + \mu(t - t_1)$$

$$E\left[X_t^2\right] = E\left[E\left[X_t^2 | X_{t_1}\right]\right] = E\left[\sigma^2(t - t_1) + (X_{t_1} + \mu(t - t_1))^2\right] = \sigma^2(t - t_1) + E[X_{t_1}^2] + \mu^2(t - t_1)^2 + 2E[X_{t_1}]\mu(t - t_1)$$

$$\mathrm{Var}[X_t] = E\left[X_t^2\right] - \left(E\left[X_t\right]\right)^2 = \sigma^2(t - t_1) + E[X_{t_1}^2] - \left(E\left[X_{t_1}\right]\right)^2 = \sigma^2(t - t_1) + \mathrm{Var}[X(t_1)].$$

Observemos que la tendencia y la varianza, a partir del primer instante de observación, dependen de los momentos de X_{t_1}. En este caso concreto, y dado que hemos supuesto una distribución normal en t_1, con media η y varianza δ^2, entonces

$$E\left[X_t\right] = \eta + \mu(t - t_1)$$

$$\mathrm{Var}[X_t] = \delta^2 + \sigma^2(t - t_1)$$

por lo que, a partir de una muestra observada, sus estimaciones máximo verosímiles son

$$\widehat{E\left[X_t\right]} = \widehat{\eta} + \widehat{\mu}(t - t_1)$$

$$\widehat{\mathrm{Var}[X_t]} = \widehat{\delta}^2 + \widehat{\sigma}^2(t - t_1)$$

Estimación máximo verosímil en el proceso de Ornstein-Uhlenbeck con factores exógenos.

El proceso de Ornstein-Uhlenbeck con factores exógenos es un proceso de difusión $\{X_t; t \geq t_0\}$ con momentos infinitesimales $A_1(x,t) = -g_{\boldsymbol{\theta}}(t)x$ y $A_2(x,t) = \sigma^2$ donde $\sigma > 0$ y $g_{\boldsymbol{\theta}}$ es una función continua dependiendo de un parámetro $\boldsymbol{\theta}$ (posiblemente multidimensional).

Estimación máximo verosímil en el proceso de Ornstein-Uhlenbeck con factores exógenos.

El proceso de Ornstein-Uhlenbeck con factores exógenos es un proceso de difusión $\{X_t; t \geq t_0\}$ con momentos infinitesimales $A_1(x,t) = -g_{\boldsymbol{\theta}}(t)x$ y $A_2(x,t) = \sigma^2$ donde $\sigma > 0$ y $g_{\boldsymbol{\theta}}$ es una función continua dependiendo de un parámetro $\boldsymbol{\theta}$ (posiblemente multidimensional).

Este proceso es transformado del proceso Wiener, siendo su densidad de transición

$$f(x,t|y,\tau) = \frac{1}{\sqrt{2\pi\,\sigma^2 e^{-2\int_\tau^t g_{\boldsymbol{\theta}}(v)\,dv} \displaystyle\int_\tau^t e^{2\int_\tau^s g_{\boldsymbol{\theta}}(v)\,dv}\,ds}} \exp\left(-\frac{\left(x - y e^{-\int_\tau^t g_{\boldsymbol{\theta}}(v)\,dv}\right)^2}{2\sigma^2 e^{-2\int_\tau^t g_{\boldsymbol{\theta}}(v)\,dv} \displaystyle\int_\tau^t e^{2\int_\tau^s g_{\boldsymbol{\theta}}(v)\,dv}\,ds} \right).$$

Estimación máximo verosímil en el proceso de Ornstein-Uhlenbeck con factores exógenos.

El proceso de Ornstein-Uhlenbeck con factores exógenos es un proceso de difusión $\{X_t; t \geq t_0\}$ con momentos infinitesimales $A_1(x,t) = -g_{\boldsymbol{\theta}}(t)x$ y $A_2(x,t) = \sigma^2$ donde $\sigma > 0$ y $g_{\boldsymbol{\theta}}$ es una función continua dependiendo de un parámetro $\boldsymbol{\theta}$ (posiblemente multidimensional).

Este proceso es transformado del proceso Wiener, siendo su densidad de transición

$$f(x,t|y,\tau) = \frac{1}{\sqrt{2\pi\,\sigma^2 e^{-2\int_\tau^t g_{\boldsymbol{\theta}}(v)\,dv} \displaystyle\int_\tau^t e^{2\int_\tau^s g_{\boldsymbol{\theta}}(v)\,dv}\,ds}} \exp\left(-\frac{\left(x - ye^{-\int_\tau^t g_{\boldsymbol{\theta}}(v)\,dv}\right)^2}{2\sigma^2 e^{-2\int_\tau^t g_{\boldsymbol{\theta}}(v)\,dv} \displaystyle\int_\tau^t e^{2\int_\tau^s g_{\boldsymbol{\theta}}(v)\,dv}\,ds}\right).$$

Observemos que en este caso la distribución de transición $X_t|X_\tau = y$ se corresponde con la de una variable normal de media

$$ye^{-\displaystyle\int_\tau^t g_{\boldsymbol{\theta}}(v)\,dv}$$

y varianza

$$\sigma^2 e^{-2\displaystyle\int_\tau^t g_{\boldsymbol{\theta}}(v)\,dv} \int_\tau^t e^{2\displaystyle\int_\tau^s g_{\boldsymbol{\theta}}(v)\,dv}\,ds$$

Estimación máximo verosímil en el proceso de Ornstein-Uhlenbeck con factores exógenos.

El proceso de Ornstein-Uhlenbeck con factores exógenos es un proceso de difusión $\{X_t; t \geq t_0\}$ con momentos infinitesimales $A_1(x,t) = -g_{\boldsymbol{\theta}}(t)x$ y $A_2(x,t) = \sigma^2$ donde $\sigma > 0$ y $g_{\boldsymbol{\theta}}$ es una función continua dependiendo de un parámetro $\boldsymbol{\theta}$ (posiblemente multidimensional).

Este proceso es transformado del proceso Wiener, siendo su densidad de transición

$$f(x,t|y,\tau) = \frac{1}{\sqrt{2\pi\,\sigma^2 e^{-2\int_\tau^t g_{\boldsymbol{\theta}}(v)\,dv} \displaystyle\int_\tau^t e^{2\int_\tau^s g_{\boldsymbol{\theta}}(v)\,dv}\,ds}} \exp\left(-\frac{\left(x - ye^{-\int_\tau^t g_{\boldsymbol{\theta}}(v)\,dv}\right)^2}{2\sigma^2 e^{-2\int_\tau^t g_{\boldsymbol{\theta}}(v)\,dv} \displaystyle\int_\tau^t e^{2\int_\tau^s g_{\boldsymbol{\theta}}(v)\,dv}\,ds}\right).$$

Consideremos ahora una muestra discreta del proceso $\mathbf{x} = (x_1, \ldots, x_n)^t$ tomada en los instantes de tiempo $\{t_1, \ldots, t_n\}$. Asimismo consideremos en este caso que $\mathrm{P}\left[X(t_1) = x_1\right] = 1$, con lo cual la verosimilitud asociada viene dada por

$$\mathbb{L}_{\mathbf{x}}(\boldsymbol{\theta}, \sigma^2) = (2\pi)^{-\frac{n-1}{2}} (\sigma^2)^{-\frac{n-1}{2}} \prod_{\alpha=2}^{n} \left[e^{-2\int_{t_{\alpha-1}}^{t_\alpha} g_{\boldsymbol{\theta}}(v)dv} \int_{t_{\alpha-1}}^{t_\alpha} e^{2\int_{t_{\alpha-1}}^{u} g_{\boldsymbol{\theta}}(v)dv} du\right]^{-\frac{1}{2}} \times$$

$$\times \exp\left(-\frac{1}{2\sigma^2} \sum_{\alpha=2}^{n} \frac{\left[x_\alpha - x_{\alpha-1} e^{-\int_{t_{\alpha-1}}^{t_\alpha} g_{\boldsymbol{\theta}}(v)dv}\right]^2}{e^{-2\int_{t_{\alpha-1}}^{t_\alpha} g_{\boldsymbol{\theta}}(v)dv} \displaystyle\int_{t_{\alpha-1}}^{t_\alpha} e^{2\int_{t_{\alpha-1}}^{u} g_{\boldsymbol{\theta}}(v)dv} du}\right)$$

Estimación máximo verosímil en el proceso de Ornstein-Uhlenbeck con factores exógenos.

El proceso de Ornstein-Uhlenbeck con factores exógenos es un proceso de difusión $\{X_t; t \geq t_0\}$ con momentos infinitesimales $A_1(x,t) = -g_{\boldsymbol{\theta}}(t)x$ y $A_2(x,t) = \sigma^2$ donde $\sigma > 0$ y $g_{\boldsymbol{\theta}}$ es una función continua dependiendo de un parámetro $\boldsymbol{\theta}$ (posiblemente multidimensional).

Este proceso es transformado del proceso Wiener, siendo su densidad de transición

$$f(x,t|y,\tau) = \frac{1}{\sqrt{2\pi\,\sigma^2 e^{-2\int_\tau^t g_{\boldsymbol{\theta}}(v)\,dv} \displaystyle\int_\tau^t e^{2\int_\tau^s g_{\boldsymbol{\theta}}(v)\,dv}\,ds}} \exp\left(-\frac{\left(x - y e^{-\int_\tau^t g_{\boldsymbol{\theta}}(v)\,dv}\right)^2}{2\sigma^2 e^{-2\int_\tau^t g_{\boldsymbol{\theta}}(v)\,dv} \displaystyle\int_\tau^t e^{2\int_\tau^s g_{\boldsymbol{\theta}}(v)\,dv}\,ds}\right).$$

Consideremos ahora una muestra discreta del proceso $\mathbf{x} = (x_1,\ldots,x_n)^t$ tomada en los instantes de tiempo $\{t_1,\ldots,t_n\}$. Asimismo consideremos en este caso que $\mathrm{P}[X(t_1) = x_1] = 1$, con lo cual la verosimilitud asociada viene dada por

$$\mathbb{L}_{\mathbf{x}}(\boldsymbol{\theta},\sigma^2) = (2\pi)^{-\frac{n-1}{2}}(\sigma^2)^{-\frac{n-1}{2}} \prod_{\alpha=2}^{n}\left[e^{-2\int_{t_{\alpha-1}}^{t_\alpha} g_{\boldsymbol{\theta}}(v)dv}\int_{t_{\alpha-1}}^{t_\alpha} e^{2\int_{t_{\alpha-1}}^{u} g_{\boldsymbol{\theta}}(v)dv}du\right]^{-\frac{1}{2}} \times$$
$$\times \exp\left(-\frac{1}{2\sigma^2}\sum_{\alpha=2}^{n}\frac{\left[x_\alpha - x_{\alpha-1}e^{-\int_{t_{\alpha-1}}^{t_\alpha} g_{\boldsymbol{\theta}}(v)dv}\right]^2}{e^{-2\int_{t_{\alpha-1}}^{t_\alpha} g_{\boldsymbol{\theta}}(v)dv}\displaystyle\int_{t_{\alpha-1}}^{t_\alpha} e^{2\int_{t_{\alpha-1}}^{u} g_{\boldsymbol{\theta}}(v)dv}du}\right)$$

Observemos cómo la forma que adopta la varianza de la transición hace prácticamente imposible abordar el problema de forma genérica. Por ello vamos a mostrar un caso concreto.

Estimación máximo verosímil en el proceso de Ornstein-Uhlenbeck con factores exógenos.

Tomemos la función $g_{\boldsymbol{\theta}}(t) = \dfrac{\theta}{t+1}$, con $\theta > -1/2$. Con ello se tiene que los momentos de la transición son

$$\begin{aligned} \mathrm{E}\left[X_t | X_s = x_s\right] &= x_s \left(\frac{s+1}{t+1}\right)^{\theta} \\ \mathrm{Var}\left[X_t | X_s = x_s\right] &= \frac{\sigma^2}{2\theta+1}\left[(t+1) - (s+1)\left(\frac{s+1}{t+1}\right)^{2\theta}\right] \end{aligned}$$

a partir de los cuales el logaritmo de la verosimilitud es

Estimación máximo verosímil en el proceso de Ornstein-Uhlenbeck con factores exógenos.

Tomemos la función $g_{\boldsymbol{\theta}}(t) = \dfrac{\theta}{t+1}$, con $\theta > -1/2$. Con ello se tiene que los momentos de la transición son

$$\begin{aligned}
\mathrm{E}\left[X_t | X_s = x_s\right] &= x_s \left(\frac{s+1}{t+1}\right)^{\theta} \\
\mathrm{Var}\left[X_t | X_s = x_s\right] &= \frac{\sigma^2}{2\theta+1}\left[(t+1) - (s+1)\left(\frac{s+1}{t+1}\right)^{2\theta}\right]
\end{aligned}$$

a partir de los cuales el logaritmo de la verosimilitud es

$$\begin{aligned}
\log\left(\mathbb{L}_{\mathbf{x}}(\theta, \sigma^2)\right) = &-\frac{n-1}{2}\log(2\pi) - \frac{n-1}{2}\log(\sigma^2) + \frac{n-1}{2}\log(2\theta+1) - \\
&-\frac{1}{2}\sum_{\alpha=2}^{n}\log\left((t_\alpha+1) - (t_{\alpha-1}+1)\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{2\theta}\right) - \\
&-\frac{2\theta+1}{2\sigma^2}\sum_{\alpha=2}^{n}\frac{\left[x_\alpha - x_{\alpha-1}\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{\theta}\right]^2}{(t_\alpha+1) - (t_{\alpha-1}+1)\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{2\theta}}
\end{aligned}$$

Estimación máximo verosímil en el proceso de Ornstein-Uhlenbeck con factores exógenos.

En cuanto a las parciales de la verosimilitud se tiene

$$\frac{\partial \log\left(\mathbb{L}_{\mathbf{x}}(\theta,\sigma^2)\right)}{\partial \sigma^2} = -\frac{n-1}{2\sigma^2} + \frac{2\theta+1}{2\sigma^4} \sum_{\alpha=2}^{n} \frac{\left[x_\alpha - x_{\alpha-1}\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{\theta}\right]^2}{(t_\alpha+1) - (t_{\alpha-1}+1)\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{2\theta}}$$

Estimación máximo verosímil en el proceso de Ornstein-Uhlenbeck con factores exógenos.

En cuanto a las parciales de la verosimilitud se tiene

$$\frac{\partial \log\left(\mathbb{L}_{\mathbf{x}}(\theta,\sigma^2)\right)}{\partial \sigma^2} = -\frac{n-1}{2\sigma^2} + \frac{2\theta+1}{2\sigma^4}\sum_{\alpha=2}^{n} \frac{\left[x_\alpha - x_{\alpha-1}\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{\theta}\right]^2}{(t_\alpha+1)-(t_{\alpha-1}+1)\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{2\theta}}$$

$$\begin{aligned}
\frac{\partial \log\left(\mathbb{L}_{\mathbf{x}}(\theta,\sigma^2)\right)}{\partial \theta} = {} & \frac{n-1}{2\theta+1} + \sum_{\alpha=2}^{n} \frac{(t_{\alpha-1}+1)\log\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{2\theta}}{(t_\alpha+1)-(t_{\alpha-1}+1)\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{2\theta}} - \\
& -\frac{1}{\sigma^2}\sum_{\alpha=2}^{n} \frac{\left[x_\alpha - x_{\alpha-1}\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{\theta}\right]^2}{(t_\alpha+1)-(t_{\alpha-1}+1)\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{2\theta}} + \\
& + \frac{2\theta+1}{\sigma^2}\sum_{\alpha=2}^{n} \frac{\log\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{\theta}\left[x_\alpha - x_{\alpha-1}\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{\theta}\right]}{\left[(t_\alpha+1)-(t_{\alpha-1}+1)\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{2\theta}\right]^2} \times \\
& \times \left[x_{\alpha-1}(t_\alpha+1) - x_\alpha(t_{\alpha-1}+1)\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{\theta}\right]
\end{aligned}$$

Estimación máximo verosímil en el proceso de Ornstein-Uhlenbeck con factores exógenos.

Planteamos las ecuaciones de verosimilitud

$$\frac{\partial \log\left(\mathbb{L}_{\mathbf{x}}(\theta,\sigma^2)\right)}{\partial\sigma^2}=0 \quad \text{y} \quad \frac{\partial \log\left(\mathbb{L}_{\mathbf{x}}(\theta,\sigma^2)\right)}{\partial\theta}=0,$$

obteniéndose, de la primera:

Estimación máximo verosímil en el proceso de Ornstein-Uhlenbeck con factores exógenos.

Planteamos las ecuaciones de verosimilitud

$$\frac{\partial \log\left(\mathbb{L}_{\mathbf{x}}(\theta,\sigma^2)\right)}{\partial \sigma^2} = 0 \quad \text{y} \quad \frac{\partial \log\left(\mathbb{L}_{\mathbf{x}}(\theta,\sigma^2)\right)}{\partial \theta} = 0,$$

obteniéndose, de la primera:

$$\widehat{\sigma}^2 = \frac{2\widehat{\theta}+1}{n-1}\sum_{\alpha=2}^{n} \frac{\left[x_\alpha - x_{\alpha-1}\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{\widehat{\theta}}\right]^2}{(t_\alpha+1)-(t_{\alpha-1}+1)\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{2\widehat{\theta}}}$$

mientras que de la segunda se tiene

Estimación máximo verosímil en el proceso de Ornstein-Uhlenbeck con factores exógenos.

Planteamos las ecuaciones de verosimilitud

$$\frac{\partial \log\left(\mathbb{L}_{\mathbf{x}}(\theta,\sigma^2)\right)}{\partial \sigma^2} = 0 \quad \text{y} \quad \frac{\partial \log\left(\mathbb{L}_{\mathbf{x}}(\theta,\sigma^2)\right)}{\partial \theta} = 0,$$

obteniéndose, de la primera:

$$\widehat{\sigma}^2 = \frac{2\widehat{\theta}+1}{n-1}\sum_{\alpha=2}^{n} \frac{\left[x_\alpha - x_{\alpha-1}\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{\widehat{\theta}}\right]^2}{(t_\alpha+1)-(t_{\alpha-1}+1)\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{2\widehat{\theta}}}$$

mientras que de la segunda se tiene

$$\sum_{\alpha=2}^{n} \frac{(t_{\alpha-1}+1)\log\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{2\widehat{\theta}}}{(t_\alpha+1)-(t_{\alpha-1}+1)\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{2\widehat{\theta}}} + \frac{2\widehat{\theta}+1}{\widehat{\sigma}^2}\sum_{\alpha=2}^{n} \frac{\log\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{\widehat{\theta}}\left[x_\alpha - x_{\alpha-1}\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{\widehat{\theta}}\right]}{\left[(t_\alpha+1)-(t_{\alpha-1}+1)\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{2\widehat{\theta}}\right]^2}$$

$$\times\left[x_{\alpha-1}(t_\alpha+1) - x_\alpha(t_{\alpha-1}+1)\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{\widehat{\theta}}\right] = 0.$$

Estimación máximo verosímil en el proceso de Ornstein-Uhlenbeck con factores exógenos.

Planteamos las ecuaciones de verosimilitud

$$\frac{\partial \log\left(\mathbb{L}_{\mathbf{x}}(\theta,\sigma^2)\right)}{\partial \sigma^2} = 0 \quad \text{y} \quad \frac{\partial \log\left(\mathbb{L}_{\mathbf{x}}(\theta,\sigma^2)\right)}{\partial \theta} = 0,$$

obteniéndose, de la primera:

$$\widehat{\sigma}^2 = \frac{2\widehat{\theta}+1}{n-1}\sum_{\alpha=2}^{n}\frac{\left[x_\alpha - x_{\alpha-1}\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{\widehat{\theta}}\right]^2}{(t_\alpha+1)-(t_{\alpha-1}+1)\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{2\widehat{\theta}}}$$

mientras que de la segunda se tiene

$$\sum_{\alpha=2}^{n}\frac{(t_{\alpha-1}+1)\log\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{2\widehat{\theta}}}{(t_\alpha+1)-(t_{\alpha-1}+1)\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{2\widehat{\theta}}} + \frac{2\widehat{\theta}+1}{\widehat{\sigma}^2}\sum_{\alpha=2}^{n}\frac{\log\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{\widehat{\theta}}\left[x_\alpha - x_{\alpha-1}\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{\widehat{\theta}}\right]}{\left[(t_\alpha+1)-(t_{\alpha-1}+1)\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{2\widehat{\theta}}\right]^2}$$
$$\times\left[x_{\alpha-1}(t_\alpha+1) - x_\alpha(t_{\alpha-1}+1)\left(\frac{t_{\alpha-1}+1}{t_\alpha+1}\right)^{\widehat{\theta}}\right] = 0.$$

Sustituyendo el valor de $\widehat{\sigma^2}$ en la segunda expresión, tenemos una ecuación bastante compleja para calcular el valor de $\widehat{\theta}$, por lo que será necesario el empleo de métodos numéricos para su resolución (dada la complejidad, obviamos el estudio de que la solución del sistema conduce al máximo, si bien es posible hacerlo en el mismo sentido que el expuesto anteriormente).