MAT-466: Modelos Lineales Generalizados

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Información

Horario:

Clases: Viernes, bloque 3-4 (09:35-10:45 hrs.) via Zoom.

Contacto:

E-mail: felipe.osorios@usm.cl.
Web: http://fosorios.mat.utfsm.cl/teaching.html y **AULA**

Evaluación:

Se realizará al menos **2 Certámenes** y **Tareas**.

Ponderaciones:

Sea $\overline{C}$ y $\overline{T}$ el promedio de certámenes y tareas, respectivamente. De este modo, la nota de presentación (NP) es dada por:

$$NP = 0.9\,\overline{C} + 0.1\,\overline{T}.$$

Criterio de aprobación:

Aquellos estudiantes que obtengan NP mayor o igual a 55 y **todos** los certámenes sobre 40, aprobarán la asignatura con nota final, $NF = NP$.

Criterio para rendir global:

En caso contrario, y siempre que $NP \geq 45$, los estudiantes podrán rendir el certamen global (CG), en cuyo caso la nota final es calculada como sigue:

$$NF = 0.6 \cdot NP + 0.4 \cdot CG.$$

Reglas adicionales

- Se llevará un control de asistencia.
- Se puede realizar preguntas sobre la materia en cualquier momento.
- Los alumnos deben apagar/silenciar sus teléfonos celulares durante clases.
- Conversaciones sobre asuntos ajenos a la clase no serán tolerados. Otros estudiantes tiene derecho a asistir clases en silencio.
- Al enviar algún e-mail al profesor, identificar el código de la asignatura en el asunto (MAT466).
- E-mail será el canal de comunicación oficial entre el profesor y los estudiantes.

Reglas: sobre los certámenes

- Es derecho del estudiante conocer la pauta de corrección la que será publicada en la página web del curso.
- Pedidos de recorrección deben ser argumentados por escrito.
- En modalidad online, Certámenes, Controles y Tareas deben ser enviados en formato **PDF**.[1]
- Cualquier tipo de fraude en prueba (copia, WhatsApp, suplantación, etc.) implicará la reprobación de los involucrados.[2]

[1] En un único archivo, orientado en una dirección legible.
[2] Puede implicar la apertura de un proceso disciplinario.

Orientaciones de estudio

- Mantener la frecuencia de estudio de inicio a final del semestre. El ideal es estudiar el contenido luego de **cada** clase.
- Estudiar primeramente el contenido dado en clases, buscando apoyo en las referencias bibliográficas.
- Las referencias son fuentes de ejemplos y ejercicios. Resuelva una buena cantidad de ejercicios. No deje esto para la víspera de la prueba.
- Buscar las referencias bibliográficas al inicio del semestre, dando preferencia a las principales y complementarias.

Programa del curso

1. Preliminares.
2. Inferencia en modelos lineales generalizados.
3. Ecuaciones de estimación generalizadas (GEE)
4. Técnicas de suavizamiento.

Bibliografía

Dobson, A. (1990).
An Introduction to Generalized Linear Models.
Chapman and Hall, London.

McCullagh, P., Nelder, J.A. (1983).
Generalized Linear Models.
Chapman and Hall, London.

Paula, G.A. (2013).
Modelos de Regressão: Com apoio computacional.
Instituto de Matemática e Estatística - Universidade de São Paulo, Brasil.

Wei, B. (1998).
Exponential Family Nonlinear Models.
Springer, Singapore.

Definición 1 (Familia exponencial):

Considere la familia de distribuciones $\mathsf{P}_{\theta,\phi}$ de un vector aleatorio $\boldsymbol{Y} = (Y_1, \dots, Y_n)^\top$ con función de densidad

$$f(\boldsymbol{y}; \boldsymbol{\theta}, \phi) = \exp[\phi\{\boldsymbol{y}^\top \boldsymbol{\theta} - b(\boldsymbol{\theta}) - s(\boldsymbol{y}, \phi)\}], \qquad (1)$$

donde $b(\cdot)$ y $s(\cdot, \cdot)$ son funciones apropiadas, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)^\top \in \Theta \subset \mathbb{R}^n$ representa el parámetro natural y $\phi > 0$ denota el parámetro de dispersión. Cuando $\boldsymbol{Y}$ tiene la densidad en (1) decimos que sigue una familia exponencial y anotamos $\boldsymbol{Y} \sim \mathsf{ED}(\boldsymbol{\theta}, \phi)$.

Observación:

Frecuentemente (1) es reescrita de forma equivalente como

$$f(\boldsymbol{y}; \boldsymbol{\theta}, \phi) = \exp[\phi\{\boldsymbol{y}^\top \boldsymbol{\theta} - b(\boldsymbol{\theta})\} + c(\boldsymbol{y}, \phi)]. \qquad (2)$$

Ejemplo:

Suponga $\boldsymbol{Y} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Tenemos que su función de densidad puede ser escrita como

$$\begin{aligned} f(\boldsymbol{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\{-\tfrac{1}{2}(\boldsymbol{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{\mu})\} \\ &= \exp\{\boldsymbol{y}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \tfrac{1}{2}\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \boldsymbol{y}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{y} - \tfrac{1}{2} \log |2\pi \boldsymbol{\Sigma}|\}, \end{aligned}$$

de este modo, el parámetro natural es $\boldsymbol{\theta} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ y $b(\boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{\theta}^\top \boldsymbol{\Sigma} \boldsymbol{\theta}$, mientras que

$$c(\boldsymbol{y}, \boldsymbol{\Phi}) = -\tfrac{1}{2}(\boldsymbol{y}^\top \boldsymbol{\Phi} \boldsymbol{y} - \log |2\pi \boldsymbol{\Phi}^{-1}|).$$

Casos particulares:

(a) Ambos y y θ escalares. En cuyo caso,

$$f(y;\theta,\phi)=\exp[\phi\{y\theta-b(\theta)\}+c(y,\phi)].$$

(b) Cuando los componentes de $\boldsymbol{Y}$ son independientes, tenemos

$$\begin{aligned} f(\boldsymbol{y};\boldsymbol{\theta},\phi) &= \exp\left[\phi\sum_{i=1}^{n}\{y_i\theta_i-b(\theta_i)\}+\sum_{i=1}^{n}c(y_i,\phi)\right] \\ &= \exp[\phi\{\boldsymbol{y}^\top\boldsymbol{\theta}-b^*(\boldsymbol{\theta})\}+c^*(\boldsymbol{y},\phi)], \end{aligned}$$

con $b^*(\boldsymbol{\theta})=\sum_{i=1}^{n}b(\theta_i)$ y $c^*(\boldsymbol{y};\phi)=\sum_{i=1}^{n}c(y_i,\phi)$.

Resultado 1 (Función generadora de momentos):

Considere $\boldsymbol{Y} \sim \mathsf{ED}(\boldsymbol{\theta}, \phi)$, entonces la función generadora de momentos (MGF) asume la forma:

$$M_Y(\boldsymbol{t}) = \mathsf{E}(e^{\boldsymbol{t}^\top \boldsymbol{Y}}) = \exp[\phi\{b(\boldsymbol{\theta} + \boldsymbol{t}/\phi) - b(\boldsymbol{\theta})\}].$$

Demostración:

En efecto,

$$\begin{aligned} M_Y(\boldsymbol{t}) &= \mathsf{E}\{\exp(\boldsymbol{t}^\top \boldsymbol{Y})\} = \int \exp(\boldsymbol{t}^\top \boldsymbol{y}) \exp[\phi\{\boldsymbol{y}^\top \boldsymbol{\theta} - b(\boldsymbol{\theta})\} + c(\boldsymbol{y}, \phi)] \,\mathsf{d}\,\boldsymbol{y} \\ &= \int \exp[\phi\{\boldsymbol{y}^\top(\boldsymbol{\theta} + \boldsymbol{t}/\phi) - b(\boldsymbol{\theta} + \boldsymbol{t}/\phi) + b(\boldsymbol{\theta} + \boldsymbol{t}/\phi) - b(\boldsymbol{\theta})\} + c(\boldsymbol{y}, \phi)] \,\mathsf{d}\,\boldsymbol{y} \\ &= \exp[\phi\{b(\boldsymbol{\theta} + \boldsymbol{t}/\phi) - b(\boldsymbol{\theta})\}] \int \exp[\phi\{\boldsymbol{y}^\top(\boldsymbol{\theta} + \boldsymbol{t}/\phi) - b(\boldsymbol{\theta} + \boldsymbol{t}/\phi)\} + c(\boldsymbol{y}, \phi)] \,\mathsf{d}\,\boldsymbol{y}, \end{aligned}$$

y el resultado sigue.

Tenemos que

$$\mathsf{E}(\boldsymbol{Y}) = \frac{\partial b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \boldsymbol{\mu}, \qquad \mathsf{Cov}(\boldsymbol{Y}) = \phi^{-1}\,\frac{\partial^2 b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \phi^{-1}\boldsymbol{V}(\boldsymbol{\mu}).$$

Evidentemente,

$$\begin{aligned}
\mathsf{d}_t\, M_Y(\boldsymbol{t}) &= \exp[\phi\{b(\boldsymbol{\theta} + \boldsymbol{t}/\phi) - b(\boldsymbol{\theta})\}]\,\mathsf{d}_t[\phi\{b(\boldsymbol{\theta} + \boldsymbol{t}/\phi) - b(\boldsymbol{\theta})\}] \\
&= \exp[\phi\{b(\boldsymbol{\theta} + \boldsymbol{t}/\phi) - b(\boldsymbol{\theta})\}]\,\phi\,\dot{b}(\boldsymbol{\theta} + \boldsymbol{t}/\phi)\,\mathsf{d}_t(\boldsymbol{\theta} + \boldsymbol{t}/\phi) \\
&= \exp[\phi\{b(\boldsymbol{\theta} + \boldsymbol{t}/\phi) - b(\boldsymbol{\theta})\}]\,\dot{b}(\boldsymbol{\theta} + \boldsymbol{t}/\phi)\,\mathsf{d}\,\boldsymbol{t}.
\end{aligned}$$

De ahí que

$$\mathsf{d}_t\, M_Y(\boldsymbol{t})\big|_{t=0} = \dot{b}(\boldsymbol{\theta})\,\mathsf{d}\,\boldsymbol{t},$$

por tanto

$$\mathsf{E}(\boldsymbol{Y}) = \dot{b}(\boldsymbol{\theta}) = \frac{\partial b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}.$$

Por otro lado,

$$\begin{aligned}\mathsf{d}_t^2\, M_Y(\boldsymbol{t}) &= \mathsf{d}_t\, M_Y(\boldsymbol{t})\, \dot{b}(\boldsymbol{\theta} + \boldsymbol{t}/\phi)\, \mathsf{d}\,\boldsymbol{t} + M_Y(\boldsymbol{t})\, \mathsf{d}\, \dot{b}(\boldsymbol{\theta} + \boldsymbol{t}/\phi)\, \mathsf{d}\,\boldsymbol{t} \\ &= (\mathsf{d}\,\boldsymbol{t})^\top M_Y(\boldsymbol{t})\, \dot{b}(\boldsymbol{\theta} + \boldsymbol{t}/\phi)\dot{b}^\top(\boldsymbol{\theta} + \boldsymbol{t}/\phi)\, \mathsf{d}\,\boldsymbol{t} + \phi^{-1} M_Y(\boldsymbol{t})(\mathsf{d}\,\boldsymbol{t})^\top \ddot{b}(\boldsymbol{\theta} + \boldsymbol{t}/\phi)\, \mathsf{d}\,\boldsymbol{t}.\end{aligned}$$

Evaluando en $\boldsymbol{t} = \boldsymbol{0}$, sigue que

$$\mathsf{d}_t^2\, M_Y(\boldsymbol{t})\big|_{t=0} = (\mathsf{d}\,\boldsymbol{t})^\top \{\dot{b}(\boldsymbol{\theta})\dot{b}^\top(\boldsymbol{\theta}) + \phi^{-1}\ddot{b}(\boldsymbol{\theta})\}\, \mathsf{d}\,\boldsymbol{t}.$$

Así

$$\mathsf{E}(\boldsymbol{Y}\boldsymbol{Y}^\top) = \dot{b}(\boldsymbol{\theta})\dot{b}^\top(\boldsymbol{\theta}) + \phi^{-1}\ddot{b}(\boldsymbol{\theta}),$$

luego

$$\mathsf{Cov}(\boldsymbol{Y}) = \mathsf{E}(\boldsymbol{Y}\boldsymbol{Y}^\top) - \mathsf{E}(\boldsymbol{Y})\,\mathsf{E}^\top(\boldsymbol{Y}) = \phi^{-1}\ddot{b}(\boldsymbol{\theta}) = \phi^{-1}\boldsymbol{V}(\boldsymbol{\mu}).$$

Observación:

Debemos resaltar que la función de varianza $\boldsymbol{V}(\boldsymbol{\mu})$ para el modelo en (1) o (2), es matriz diagonal.

Lema 1:

Considere $\boldsymbol{Y} \sim \text{ED}(\boldsymbol{\theta}, \phi)$, con función de densidad (2). Entonces,

$$\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}^\top} = \boldsymbol{V}(\boldsymbol{\mu}), \qquad \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\mu}^\top} = \boldsymbol{V}^{-1}(\boldsymbol{\mu}).$$

Demostración:

Como $\boldsymbol{\mu} = \dot{b}(\boldsymbol{\theta})$ tenemos $\boldsymbol{\theta} = \dot{b}^{-1}(\boldsymbol{\mu})$, donde $\dot{b}^{-1}(\cdot)$ es la función inversa de $\dot{b}(\cdot)$. Por otro lado, la función de varianza

$$\ddot{b}(\boldsymbol{\theta}) = \ddot{b}(\dot{b}^{-1}(\boldsymbol{\mu})) = \boldsymbol{V}(\boldsymbol{\mu}).$$

De ahí que $\partial \boldsymbol{\mu}/\partial \boldsymbol{\theta}^\top = \ddot{b}(\boldsymbol{\theta}) = \boldsymbol{V}(\boldsymbol{\mu})$. Note también que

$$\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\mu}^\top} = \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}^\top} \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\mu}^\top} = \boldsymbol{I}_n.$$

lo que lleva a

$$\boldsymbol{V}(\boldsymbol{\mu}) \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\mu}^\top} = \boldsymbol{I}_n \qquad \Rightarrow \qquad \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\mu}^\top} = \boldsymbol{V}^{-1}(\boldsymbol{\mu}).$$

Familia exponencial

Sea $\boldsymbol{Y} \sim \mathsf{ED}(\boldsymbol{\theta}, \phi)$, entonces la función de log-verosimilitud adopta la forma:

$$\ell(\boldsymbol{\mu}; \boldsymbol{y}) = \phi\{\boldsymbol{y}^\top \boldsymbol{\theta} - b(\boldsymbol{\theta})\} + c(\boldsymbol{y}, \boldsymbol{\theta}).$$

Las ecuaciones de verosimilitud resultan

$$\boldsymbol{U}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\mu}; \boldsymbol{y})}{\partial \boldsymbol{\theta}} = \phi(\boldsymbol{Y} - \boldsymbol{\mu})$$
$$\boldsymbol{U}(\boldsymbol{\mu}) = \frac{\partial \ell(\boldsymbol{\mu}; \boldsymbol{y})}{\partial \boldsymbol{\mu}} = \phi \boldsymbol{V}^{-1}(\boldsymbol{\mu})(\boldsymbol{Y} - \boldsymbol{\mu}).$$

Es fácil ver que el MLE de $\boldsymbol{\mu}$ es $\widehat{\boldsymbol{\mu}} = \boldsymbol{Y}$ y $\widehat{\boldsymbol{\theta}} = \dot{b}^{-1}(\widehat{\boldsymbol{\mu}})$. Tenemos también,

$$U(\phi) = \frac{\partial \ell(\boldsymbol{\mu}; \boldsymbol{y})}{\partial \phi} = \{\boldsymbol{y}^\top \boldsymbol{\theta} - b(\boldsymbol{\theta})\} + \dot{c}(\boldsymbol{y}; \phi), \qquad \frac{\partial^2 \ell(\boldsymbol{\mu}; \boldsymbol{y})}{\partial \phi^2} = \ddot{c}(\boldsymbol{y}; \phi).$$

Además,

$$\frac{\partial \ell(\boldsymbol{\mu}; \boldsymbol{y})}{\partial \phi \partial \boldsymbol{\theta}} = \boldsymbol{Y} - \boldsymbol{\mu} \qquad \frac{\partial \ell(\boldsymbol{\mu}; \boldsymbol{y})}{\partial \phi \partial \boldsymbol{\mu}} = \boldsymbol{V}^{-1}(\boldsymbol{\mu})(\boldsymbol{Y} - \boldsymbol{\mu}),$$

de donde obtenemos $\mathsf{E}(-\ddot{\ell}_{\phi\theta}) = \boldsymbol{0}$ y $\mathsf{E}(-\ddot{\ell}_{\phi\mu}) = \boldsymbol{0}$.

Suponga $\boldsymbol{Y} \sim \text{ED}(\boldsymbol{\theta}, \phi)$ con ϕ fijo y considere la hipótesis,

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0, \quad \text{versus} \quad H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0.$$

El estadístico de razón de verosimilitudes adopta la forma:

$$\begin{aligned}
LR &= 2\{\ell(\widehat{\boldsymbol{\mu}}; \boldsymbol{y}) - \ell(\boldsymbol{\mu}_0; \boldsymbol{y})\} \\
&= 2[\phi\{\boldsymbol{y}^\top \boldsymbol{\theta} - b(\boldsymbol{\theta})\}_{\mu=y} - \phi\{\boldsymbol{y}^\top \boldsymbol{\theta} - b(\boldsymbol{\theta})\}_{\mu=\mu_0}] \\
&= 2\phi[\{\boldsymbol{y}^\top \boldsymbol{h}(\boldsymbol{y}) - b(\boldsymbol{h}(\boldsymbol{y}))\} - \{\boldsymbol{y}^\top \boldsymbol{h}(\boldsymbol{\mu}_0) - b(\boldsymbol{h}(\boldsymbol{\mu}_0))\}],
\end{aligned}$$

donde $\boldsymbol{\theta} = \dot{b}^{-1}(\boldsymbol{\mu}) = \boldsymbol{h}(\boldsymbol{\mu})$. Es bien conocido que $LR \xrightarrow{\text{D}} \chi^2(n)$. De este modo,

$$LR(\boldsymbol{\mu}) = D^*(\boldsymbol{y}, \boldsymbol{\mu}) = \phi D(\boldsymbol{y}, \boldsymbol{\mu}) \overset{\text{a}}{\sim} \chi^2(n),$$

donde

$$\begin{aligned}
D(\boldsymbol{y}, \boldsymbol{\mu}) &= 2\{\boldsymbol{y}^\top \boldsymbol{h}(\boldsymbol{y}) - b(\boldsymbol{h}(\boldsymbol{y}))\} - 2\{\boldsymbol{y}^\top \boldsymbol{h}(\boldsymbol{\mu}) - b(\boldsymbol{h}(\boldsymbol{\mu}))\} \\
&= 2\{\boldsymbol{y}^\top \boldsymbol{\theta} - b(\boldsymbol{\theta})\}_{\mu=y} - 2\{\boldsymbol{y}^\top \boldsymbol{\theta} - b(\boldsymbol{\theta})\},
\end{aligned}$$

es llamada función desvio (o devianza).

Note que

$$\ell(\boldsymbol{\mu}, \boldsymbol{y}) = -\frac{1}{2}\phi D(\boldsymbol{y}, \boldsymbol{\mu}) + c(\boldsymbol{y}, \phi) + \phi\{\boldsymbol{y}^\top \boldsymbol{\theta} - b(\boldsymbol{\theta})\}_{\mu=y}.$$

Para $\boldsymbol{Y} \sim \mathsf{ED}(\boldsymbol{\theta}, \phi)$ la función de log-verosimilitud puede ser representada como:

$$f(\boldsymbol{y}; \boldsymbol{\mu}, \phi) = f(\boldsymbol{y}; \boldsymbol{y}, \phi) \exp\Big\{ -\frac{1}{2}\phi D(\boldsymbol{y}, \boldsymbol{\mu})\Big\}$$

$$\ell(\boldsymbol{\mu}, \phi; \boldsymbol{y}) = \ell(\boldsymbol{y}, \phi; \boldsymbol{y}) - \frac{1}{2}\phi D(\boldsymbol{y}, \boldsymbol{\mu})$$

En efecto,

$$\begin{aligned} \frac{f(\boldsymbol{y}; \boldsymbol{\mu}, \phi)}{f(\boldsymbol{y}; \boldsymbol{y}, \phi)} &= \exp[\phi\{\boldsymbol{y}^\top \boldsymbol{\theta} - b(\boldsymbol{\theta})\} - \phi\{\boldsymbol{y}^\top \boldsymbol{\theta} - b(\boldsymbol{\theta})\}_{\mu=y}] \\ &= \exp\Big\{ -\frac{1}{2}\phi D(\boldsymbol{y}, \boldsymbol{\mu})\Big\}. \end{aligned}$$

Ejemplo:

Sea $\boldsymbol{Y} = (Y_1, \dots, Y_n)^\top$ y suponga que $\boldsymbol{Y} \sim \mathsf{N}_n(\boldsymbol{\mu}, \sigma^2\boldsymbol{\Sigma})$.[3] Sea

$$D(\boldsymbol{y}; \boldsymbol{\mu}) = (\boldsymbol{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}).$$

Entonces, podemos escribir la función de densidad como:

$$f(\boldsymbol{y}; \boldsymbol{\mu}, \sigma^2) = |2\pi\sigma^2\boldsymbol{\Sigma}|^{-1/2} \exp\Big\{ - \frac{1}{2\sigma^2} D(\boldsymbol{y}; \boldsymbol{\mu})\Big\}.$$

De este modo, la función de log-verosimilitud adopta la forma:

$$\ell(\boldsymbol{\mu}, \sigma^2; \boldsymbol{y}) = -\frac{1}{2}\log|2\pi\boldsymbol{\Sigma}| - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}D(\boldsymbol{y}; \boldsymbol{\mu}).$$

Así,

$$\boldsymbol{U}(\boldsymbol{\mu}) = -\frac{1}{2\sigma^2}\frac{\partial}{\partial\boldsymbol{\mu}}D(\boldsymbol{y}; \boldsymbol{\mu}) = \frac{1}{\sigma^2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})$$

$$\boldsymbol{\mathcal{F}}(\boldsymbol{\mu}) = \mathsf{E}\Big\{ - \frac{\partial^2\ell(\boldsymbol{\mu}, \sigma^2; \boldsymbol{y})}{\partial\boldsymbol{\mu}\partial\boldsymbol{\mu}^\top}\Big\} = \frac{1}{\sigma^2}\boldsymbol{\Sigma}^{-1}.$$

Evidentemente el MLE de $\boldsymbol{\mu}$ es $\widehat{\boldsymbol{\mu}} = \boldsymbol{Y}$.

[3] En efecto, $\mathsf{E}(\boldsymbol{Y}) = \boldsymbol{\mu}$ y $\mathsf{Cov}(\boldsymbol{Y}) = \sigma^2\boldsymbol{\Sigma}$.