Ventajas del enfoque Bayesiano en Salud Pública

Hèctor Perpiñán
Rubén Amorós, Anabel Forte, Silvia Lladosa y Facundo Muñoz

Jornades de consultoria estadística i software II

24 de octubre de 2013

BayeStats
STATISTICAL SERVICES

Esquema

Introducción
¿Cómo aprendemos en la vida?

JORNADA 10 - Primera División Española

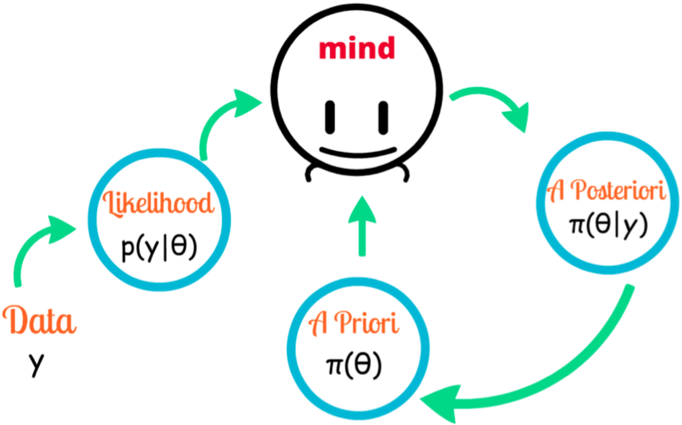
- ¿Quién ganará?

- ¿De qué información disponemos?

Datos Últimos resultados de los Barça-Madrid

¿Algo más? ¿A favor de quién estoy?...**Conocimiento subjetivo**

Introducción
¿Cómo funciona la Estadística Bayesiana?



- Procesos de aprendizaje del ser humano
- Los médicos incorporan cada nuevo paciente (datos) a su conocimiento (información a priori) para usarlo con el próximo paciente (información a posteriori)
- Toma de decisiones

¿Siempre debemos incluir información previa?

$$\pi(\theta|y) = \frac{p(y|\theta)\pi(\theta)}{\pi(y)} \propto p(y|\theta)\,\pi(\theta)$$

- Información a posteriori
- Datos
- Información a priori

Diferentes tipos de previas:

- Objetivas
- Subjetivas

Interpretación

- Resultados en función de la distribución a posteriori
- Estima P(Hipótesis|Datos) en lugar de P(Datos|Hipótesis)
- Las probabilidades son fáciles de entender

- La distribución a posteriori habitualmente no se puede calcular analíticamente.

Existen dos alternativas prácticas para su cómputo:

Markov-Chain Monte Carlo Obtienen una *muestra* simulada de la distribución a posteriori.

Métodos aproximados Utilizan métodos numéricos de integración y otras aproximaciones para el cálculo de la distribución a posteriori.

(Win/Open)BUGS (Bayesian inference Using Gibbs Sampling). 1990.
Primer paquete MCMC de propósito general, responsable en buena medida de la popularización de la estadística Bayesiana.

JAGS (Just Another Gibbs Sampler). 2000.
Una modernización de BUGS multiplataforma y extensible.

STAN (Hamiltonian Monte Carlo). 2010.
Implementa un algoritmo MCMC novedoso y más eficiente.

OTROS

Encontramos paquetes de `R` (`MCMCpack`; `lmm`; `MCMCglmm`; `spBayes`; ...), y otros que funcionan de manera autónoma (`EpiDat`; `BayesX`; ...).

INLA (Integrated Nested Laplace Approximation). 2009. Permite ajustar una familia muy amplia de modelos estadísticos (*latent Gaussian models*) sin utilizar simulación (i.e. en segundos o minutos).

¿Qué permite hacer?

- GLM, GLMM, GAM, GAMM
- Series temporales
- Modelos longitudinales
- Modelos espacio-temporales
- Análisis de supervivencia
- Modelos dinámicos
- Regresión semiparamétrica
- Log-Gaussian Cox-processes
- etc.

Existe un paquete en `R` que proporciona una interfaz amigable.

Ejemplo
Introducción

Estudio realizado para Salvador Marí (Proyecto doctoral).

Medidas antropométricas y valor nutricional en niños con autismo

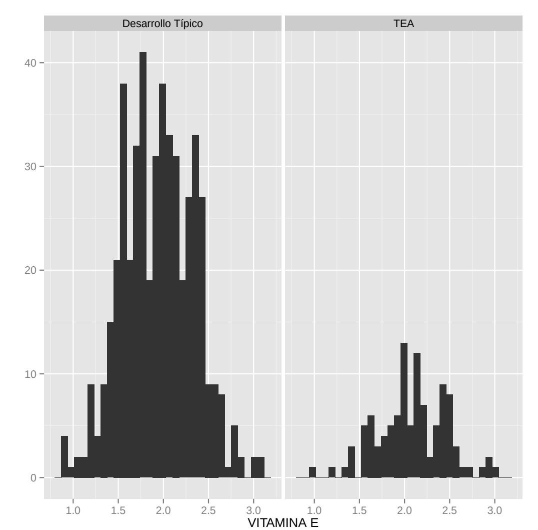
- Datos tomados en el Área de Valencia, concretamente en todos aquellos Centros de Educación Especial (CEE) y aulas CYL (Comunicación y Lenguaje) de colegios de primaria con niños diagnosticados con Transtornos del Espectro Autista (TEA).
- Niños de entre 6 y 9 años.

OBJETIVO 1: Examinar el estado nutricional en niños diagnosticados con TEA y compararlo con niños con desarrollo típico, en base a una serie de covariables.

OBJETIVO 2: Determinar las probabilidades de cumplir los criterios nutricionales marcados por la UE (para todos los nutrientes) según una serie de covariables.

Ejemplo
OBJETIVO 1: Regresión lineal

Variable Respuesta: VITAMINA E —> log(VITAMINA E)



Covariables

- Covariables discretas:

1. Autista
2. Sexo
3. Dieta restrictiva (sin gluten, sin lactosa, etc.)

- Covariables continuas:

1. Edad
2. Peso
3. Talla

Ejemplo
OBJETIVO 1: Regresión lineal

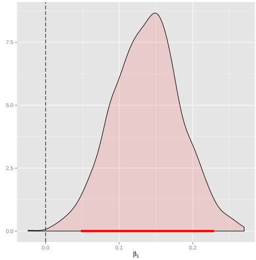
Para cada niño i, $i = 1, ..., 600$ consideramos:

$$y_i = \mathbf{X}_i' \boldsymbol{\beta} + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2)$$

Parám.	Frecuentista			Bayesiano		
	Valor	Int. Conf. 95 %	P-valor	Media	Int. Cred. 95 %	P(β>0)
β_0	0.490	(-0.075, 1.054)	0.045	0.493	(-0.100, 1.054)	0.953
β_1 (TEA)	0.138	(0.048, 0.229)	0.001	0.137	(0.047, 0.226)	0.999
β_2 (sexo)	0.027	(-0.038, 0.092)	0.206	0.027	(-0.034, 0.091)	0.795
β_3 (dieta)	0.112	(-0.064, 0.287)	0.107	0.108	(-0.066, 0.275)	0.876
β_4 (edad)	-0.001	(-0.004, 0.002)	0.296	-0.001	(-0.004, 0.002)	0.287
β_5 (peso)	0.003	(-0.004, 0.009)	0.221	0.003	(-0.004, 0.009)	0.771
β_6 (talla)	0.011	(0.005, 0.018)	0.000	0.011	(0.005, 0.018)	1.000

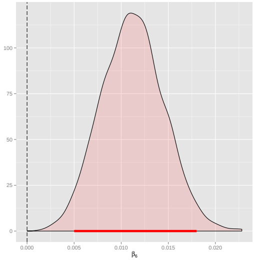
Ejemplo
OBJETIVO 1: Regresión lineal

¿Tienen relación nuestras covariables con la variable repuesta?



P-valor=0.001
P(β_1>0)=0.999

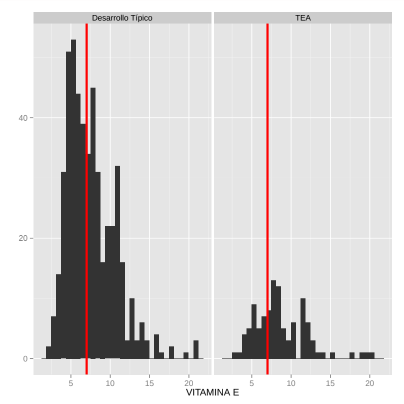
P-valor=0.107
P(β_3>0)=0.876



P-valor=0.000
P(β_6>0)=1.000

Ejemplo
OBJETIVO 2: Regresión logística

Variable Respuesta: VITAMINA E ≥ 7



Covariables

Mantenemos las mismas covariables.

Ejemplo

OBJETIVO 2: Regresión logística

Para cada niño i, $i = 1, ..., 600$ consideramos:

$$y_i \sim Bern(p_i)$$

$$logit(p_i) = \mathbf{X}_i' \boldsymbol{\beta}$$

Parám.	Frecuentista			Bayesiano		
	Valor	Int. Conf. 95 %	P-valor	Media	Int. Cred. 95 %	P(β>0)
β_0	-6.556	(-9.819,-3.384)	0.000	-6.707	(-9.819,-3.267)	0.001
β_1 (TEA)	0.717	(0.211, 1.241)	0.003	0.738	(0.220, 1.302)	0.997
β_2 (sexo)	0.204	(-0.151, 0.560)	0.130	0.214	(-0.148, 0.569)	0.873
β_3 (dieta)	0.686	(-0.336, 1.815)	0.102	0.750	(-0.281, 1.773)	0.918
β_4 (edad)	0.007	(-0.011, 0.024)	0.232	0.007	(-0.011, 0.026)	0.784
β_5 (peso)	0.020	(-0.018, 0.058)	0.154	0.020	(-0.017, 0.060)	0.843
β_6 (talla)	0.041	(0.005, 0.077)	0.014	0.042	(0.004, 0.075)	0.985

Ejemplo

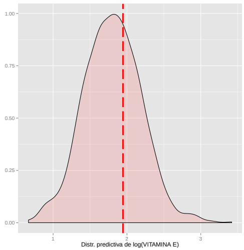
OBJETIVO 2: Regresión logística vs. Regresión lineal bayesiana

Sea
$\tilde{x} = (\text{NO Autista},\ \text{Niño},\ \text{NO Dieta},\ \text{Edad} = 72,\ \text{Peso} = 20,\ \text{Talla} = 115)$

	Frecuentista	Bayesiano
p_i	0.309	0.306

Podemos calcular la distribución predictiva para unos nuevos datos $\tilde{y}$ como:

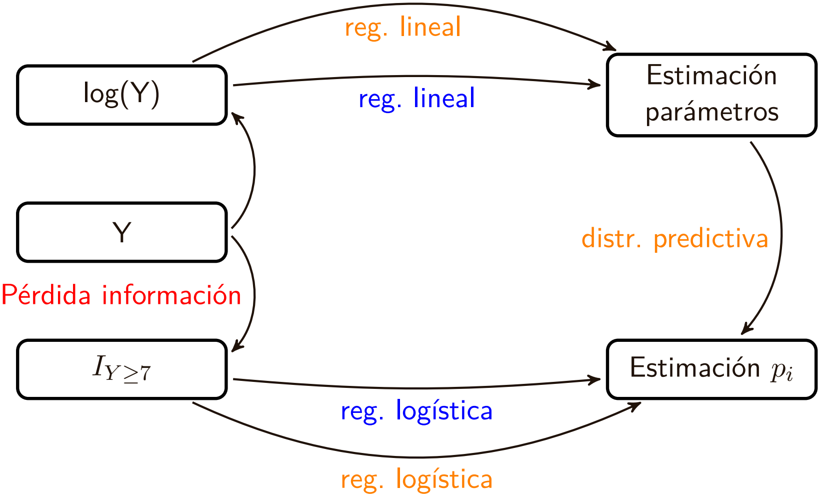
$$p(\tilde{y}|y) = \int p(\tilde{y}|\theta)\pi(\theta|y)d\theta$$



$$p_i = p(\tilde{y} \geq log(7)|y) = 0.353$$

Ejemplo
Conclusiones

FRECUENTISTA-BAYESIANO



MUCHAS GRACIAS POR LA ATENCIÓN!!!