Universidad de Oviedo

Modelos bayesianos. Enero de 2025

Ejercicio 1 (3 puntos). Un grupo de control de 30 ratas bebe agua normal mientras otro grupo de 30 ratas consume agua contaminada con una dosis diaria de 30 mg/kg de perclorato de amonio. A las 19 semanas se examinan todas las ratas y se detectan 0 y 2 ratas con tumor de tiroides en el grupo control y en el de experimentación, respectivamente. En 67 estudios previos del Programa Nacional de Toxicología de un total de 3 919 ratas de control con dos años de vida se han detectado 38 con tumores tiroidales y se estima que la proporción de que estas ratas enfermaran en las primeras 19 semanas fue de 0.0005.

Sea θ (`theta`) la probabilidad de contraer el tumor en las 19 primeras semanas en ratas sanas, γ (`gamma`) la probabilidad de contraer tumor en dos años de vida y ρ (`rho`) la proporción de tumores que se han desarrollado en los primeras 19 semanas. El código de R para resolver estadísticamente este problema se presenta a continuación. Justifique adecuadamente las respuestas.

1. (1 pto.) Interprete y contextualice el código de programación empleado para analizar el experimento:
   a) `prop.test(x=c(0,2),n=c(30,30),alternative="less")`,
   b) `rbeta(n,38,3881)`,
   c) `rbinom(n,30,theta)`,
   d) `mean(y_a>=2)`.
2. (0.5 ptos.) Detalle un test estadístico frecuentista apropiado al problema y determine qué concluye de este test. Interprete el significado del p-valor.
3. (0.5 ptos.) ¿Cuál es la probabilidad de observar 2 o más ratas con tumor en el grupo de tratamiento suponiendo la hipótesis nula de que el perclorato es inocuo?
4. (1 pto.) ¿Qué conclusiones saca de este estudio sobre la incidencia del perclorato de amonio en el desarrollo de tumores tiroidales?

```
prop.test(x=c(0,2),n=c(30,30),alternative="less")
##  2-sample test for equality of proportions \
## with continuity correction
## data:  c(0, 2) out of c(30, 30)
## X-squared = 0.51724, df = 1, p-value = 0.236
## alternative hypothesis: less
## 95 percent confidence interval:
##  -1.0000000  0.0415766
## sample estimates:
##     prop 1     prop 2
## 0.00000000 0.06666667

set.seed(67)
n <- 1e6
gamma <- rbeta(n,38,3881)
rho <- rbeta(n,0.11,2.6) ## Median: 0.0005
theta  <- gamma*rho
y <- rbinom(n,30,theta)
theta_a  <- theta[y==0]
y_a <- rbinom(length(theta_a),30,theta_a)
mean(y_a>=2) ## (a) 0.0004450378
theta_b  <- theta[y>=2]
y_b <- rbinom(length(theta_b),30,theta_b)
mean(y_b>=2) ## (b) 0.0101833
theta_c  <- theta[y==2]
y_c <- rbinom(length(theta_c),30,theta_c)
mean(y_c>=2) ## (c) 0.004201681
```

Ejercicio 2 (1 punto). Reflexione sobre el siguiente texto y proponga argumentos a favor o en contra sobre las ideas que presenta.

> «Parménides y Heráclito debaten sobre la altura media de

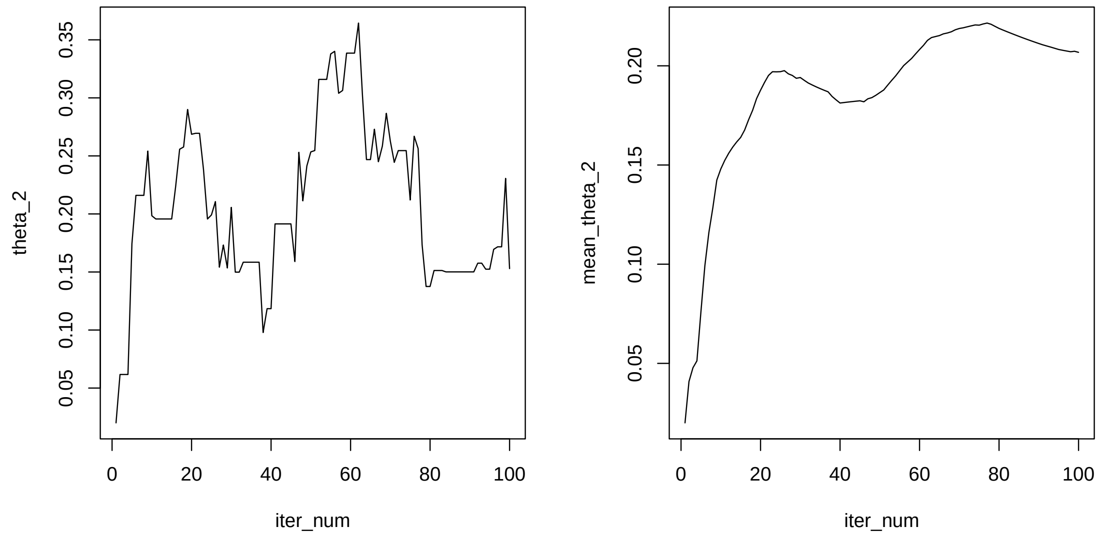
los españoles μ en un momento dado:

$$\begin{cases} H_0 : \mu = 162, \\ H_1 : \mu \neq 162. \end{cases}$$

»Parménides, filósofo griego del siglo V a.C., argumenta que la realidad es una, inmutable e indestructible, el cambio y el movimiento son ilusiones mientras que la verdadera realidad es estática y eterna. Imagine un bloque de mármol: para Parménides este mármol no cambia, no se mueve. Si lo tallamos, la esencia del mármol permanece intacta. Lo que percibimos como cambio es solo una ilusión de nuestros sentidos, sujetos a errores y confusiones. La verdadera naturaleza del mármol es inmutable y eterna.

»Por lo tanto, según esta visión filosófica, la altura media real de los españoles en un momento dado (el parámetro poblacional μ) es constante y fija, aunque no la conozcamos exactamente, y las variaciones que observamos en diferentes muestras serían ilusiones circunstanciales, meras apariencias debido a la limitación de nuestros sentidos y métodos de medición. El intervalo de confianza, entonces, intenta acercarse a esa verdad inmutable, reconociendo que nuestras observaciones pueden contener errores y variaciones. [...]» Emilio Torres Manzanera, *¿Parménides o Heráclito? Estadística frecuentista versus bayesiana*, 2024, https://torres.epv.uniovi.es/.

Ejercicio 3 (2 puntos). Se disponen los siguientes gráficos de una simulación de la distribución a posteriori del parámetro θ_2 mediante una cadena de Markov:

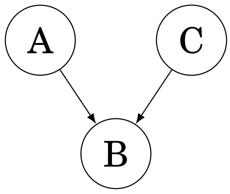


1. Argumente qué método(s) MCMC, de los vistos en clase, ha(n) podido simular dicha cadena.
2. Para calcular la distribución a posteriori, ¿cuál de los dos gráficos resulta más apropiado? ¿por qué?
3. Si se desea calcular un intervalo de credibilidad al 95%, ¿cuál de los dos gráficos resulta más apropiado? ¿resultaría fiable este intervalo?
4. Se observa en la gráfica de la derecha una línea más o menos creciente. ¿Qué conclusiones saca de este comportamiento?

Ejercicio 4 (2 puntos). Explique el modelo de regresión logística bayesiana utilizando el lenguaje matemático adecuado.

1. (0.5 ptos.) ¿Qué diferencias fundamentales encuentra respecto a los modelos bayesianos de regresión lineal?
2. (0.5 ptos.) ¿Qué es la función de enlace y qué tiene que cumplir? ¿sabría dar algún ejemplo?
3. (1 pto.) Comente cómo se actualizan las creencias sobre los parámetros al observar los datos en la práctica.

Ejercicio 5.a (1 punto). Considere la red bayesiana discreta de tres nodos de efecto común.



Obtenga una fórmula para los siguientes supuestos:

1. Observado A, calcula $Bel(B = b)$ y $Bel(C = c)$.
2. Observado B, calcula $Bel(A = a)$.
3. Observado A y B, calcula $Bel(C = c)$.

Ejercicio 5.b (1 punto). En una red bayesiana que representa las causas del cáncer de pulmón, se considera que la contaminación (P) y ser fumador (S) influyen en tener cáncer (C). Se ha considerado que la contaminación alcanza niveles altos (H) o bajos (L), mientras que ser fumador o tener cáncer puede resultar verdadero (T) o falso (F) con las siguientes distribuciones probabilísticas:

$p(P = L)$	0.85
$p(S = T)$	0.25
$p(C = T \mid P = H, S = T)$	0.06
$p(C = T \mid P = H, S = F)$	0.03
$p(C = T \mid P = L, S = T)$	0.04
$p(C = T \mid P = L, S = F)$	0.003

1. Supóngase que una persona es fumadora, ¿qué probabilidades habría de que padeciese cáncer?
2. Y partiendo de una persona diagnosticada con cáncer de pulmón, ¿cuál es la probabilidad que sea fumador?
3. Supóngase que, además de padecer cáncer, sabe que la persona ha estado expuesta a contaminación alta. ¿Cuál es la probabilidad de ser fumador?