The punctual hypothesis a Bayesian perspective

Miguel A. Gómez Villegas
Universidad Complutense
(11 de noviembre de 2011)

Grupo de Métodos Bayesianos

- Eusebio Gómez Sánchez-Manzano: Univ. Complutense
- Beatriz González-Pérez: Univ. Complutense
- José A. López Varona: Univ. Complutense
- Paloma Maín Yaque: Univ. Complutense
- Juan M. Marín Diazaraque: Univ. Carlos III
- Hilario Navarro Veguillas: Univ. a Distancia
- Mayte Rodriguez Bernal: Univ. Complutense
- Eva Romero Ramos: Univ. Europea de Madrid
- María del Rosario Susi: Univ. Complutense
- Isabel Salazar Muñoz: Univ. Complutense
- Luis Sanz Sanmiguel: Univ. Complutense
- Paola Viviani: Pontificia Univ. Católica de Chile

Summary

- P-value or frequentist evidence
- The posterior probability or Bayesian evidence
- The punctual hypothesis problem
- Contributions
- Our procedure
- Theoretical justification
- Coherence (some times) between frequentists and Bayesians
- Direct comparison
- Application to contingence tables
- Last comments

P-value or frequentist evidence

$H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$

If is observed $\vec{x} = \left(x_1, \cdots, x_n\right)$ and the rejection region of the test is

$$RC = \left\{T\left(\vec{x}\right) \geq c_\alpha\right\}$$

and c_α is the constant of the significance level α

$$\sup_{\theta \in \Theta_0} P\left\{RC \mid \theta\right\} \leq \alpha$$

Then the p-value is given by

$$p(\vec{x}) = \sup_{\theta \in \Theta_0} P\left\{T\left(\vec{X}\right) \geq T\left(\vec{x}\right) \mid \theta\right\}$$

- It is the lowest value α to which the sample point obtained belongs to rejection region

- If the size is $\alpha < p(\vec{x})$ then $T\left(\vec{x}\right) \notin RC$
- If the size is $\alpha > p(\vec{x})$ then $T(\vec{x}) \in RC$

- The prosecutor fallacy

  Lindley, D.V.(2000) The philosophy of statistics. *The Statistician,* **3,** 293-337.

Calibration of the p-value

- p-value $\in [0, 0'01]$ decisive evidence against H_o
- p-value $\in [0'01, 0'05]$ strong evidence against H_o
- p-value $\in [0'05, 0'1]$ substantial evidence against H_o
- p-value $\in [0'1, 1]$ favourable evidence to H_o

Fisher, R.A. (1956) *Statistical Method and Scientific Inference.* Edimburgo: Oliver and Boyd.

The posterior probability or Bayesian evidence

$H_0 : \theta \in \Theta_0$ frente a $H_1 : \theta \in \Theta_1$

Si se observa $\vec{x} = (x_1, \cdots, x_n)$ y se calculan

$$P(\Theta_0 \mid \vec{x}) \qquad\qquad P(\Theta_1 \mid \vec{x})$$

Si $P(\Theta_0 \mid \vec{x}) > P(\Theta_1 \mid \vec{x})$ se acepta H_0

Si $P(\Theta_0 \mid \vec{x}) < P(\Theta_1 \mid \vec{x})$ se acepta H_1

Bayes factor

$$H_0 : \theta \in \Theta_0 \quad \text{frente a} \quad H_1 : \theta \in \Theta_1$$

$$B(\vec{x}; H_0, H_1) = \frac{P(H_0 \mid \vec{x}) \big/ P(H_1 \mid \vec{x})}{P(H_0) \big/ P(H_1)} = P(\vec{x} \mid H_0) \big/ P(\vec{x} \mid H_1)$$

Calibration of the Bayes factor

$$B(\vec{x};H_0,H_1)\in\left[0,0'01\right] \text{ evidencia decisiva contra } H_0$$

$$B(\vec{x};H_0,H_1)\in\left(0'01,0'1\right] \text{ evidencia fuerte contra } H_0$$

$$B(\vec{x};H_0,H_1)\in\left(0'1,0'31\right] \text{ evidencia sustancial contra } H_0$$

$$B(\vec{x};H_0,H_1)\in\left(0'31,\infty\right) \text{ evidencia a favor de } H_0$$

- Jeffreys, H. (1939-1961) *Theory of Probability.* Oxford: University Press. Pág. 432 de la edición de 1961.

- Casella, G. and Berger, R.L. (1987) Reconciling Bayesian and frequentist evidence in the one-sided testing problem. *J. American Stat. Assoc.*, **82,** 106-111.

Reconcilian la evidencia bayesiana y frecuentista

$$H_o : \theta \geq \theta_0 \quad \text{frente a} \quad H_1 : \theta < \theta_0$$

$$\inf_{\pi \in \Gamma} P\left\{H_0 \mid \vec{x}\right\} \leq p(\vec{x})$$

Para todas las distribuciones simétricas alrededor de θ_0
Para todas las unimodales y simétricas
Para todas las normales.

Cuando el modelo es de localización, $f(x \mid \theta) = f(x - \theta)$
$f(\cdot)$ es simétrica alrededor de θ_0
$f(\cdot)$ tiene cociente de verosimilitud monótono

- Berger, J. and Sellke, T. (1987) Testing a point null hypothesis: the irreconcilability of p-values and evidence. *J. American Stat. Assoc.*, **82,** 112-122.

$$H_o : \theta = \theta_0 \quad \text{frente a} \quad H_1 : \theta \neq \theta_0$$

Encuentran disparidad entre la evidencia bayesiana y frecuentista cuando se emplea la distribución inicial

$$P^*(A) = \pi_0 I_A(\theta_0) + (1-\pi_0)\int_A \pi(\theta)d\theta \qquad \text{con} \qquad A \in \mathrm{B}_{\Theta}$$

entendida $\inf_{\pi \in \Gamma} P\{H_0 \mid \vec{x}\} \square \; p(\vec{x})$ cuando el modelo es de localización $f(x \mid \theta) = f(x - \theta)$

$$\pi(\theta) \in \Gamma_A, \Gamma_S, \Gamma_{US}, \Gamma_{NOR} \text{ y } \pi_0 = 0'5$$

$f(\cdot)$ es simétrica alrededor de θ_0

$f(\cdot)$ tiene cociente de verosimilitud monótono

En relación con este problema y entre autores españoles

- Gómez-Villegas, M.A.and Sanz, L. (1998) Reconciling Bayesian and frequentist evidence in the point null testing problem.*Test,* **7**, 207-216.
- Bayarri, M.J. and Berger, J.O. (2000) P-values for composite null models. *J. Amer. Statist. Assoc.* **95**,1127-1142.
- Gómez-Villegas, M.A.and Sanz, L. (2000) Epsilon-contaminated priors in testing point null hypothesis: a procedure to determine the prior probability.*Statist. Probab. Lett.* **47,** 53-60.
- Gómez-Villegas, M.A., Maín, P. and Sanz, L. (2002) A suitable Bayesian approach in testing point null hypothesis: some examples revisited. *Comm. Statist.-Theory Methods,* **31,** 201-217.
- De la Horra, J. and Rodríguez-Bernal, M. (2003) Bayesian robustness of the posterior predictive p-value. *Comm. Statist.-Theory Methods,* **32,** 1493-1503.
- Gómez-Villegas, M.A., Maín, P., Navarro, H. and Sanz, L. (2004) Asymptotic relationships between posterior probabilities and p-values using the hazard rate. *Stat.& Probab. Lett,* **66**, 59-66.
- Gómez-Villegas, M.A. and González, B. (2005) Bayesian analysis of contingency tables. *Comm. Statist.-Theory Methods,* **34,** 1793-1754.

- Gómez-Villegas, M.A. and González, B. (2008) Epsilon-contaminated priors in contingency tables. *Test*, **17**, 1, 163-178.
- Gómez-Villegas, M.A. and González, B. (2008) The multi-variate point null testing problem: A Bayesian discussion, *Statistics and Probability Letters,* **78,** 17, 3070-3074.
- Gómez-Villegas, M.A., Maín, P. and Sanz, L. (2009) A Bayesian analysis for the multivariate point null testing problem. *Statistics,* **43,** 4, 379-391.
- Gómez-Villegas, M.A. and González-Pérez, B. (2010) r * s Tables from a Bayesian viewpoint, *Revista Matemática Complutense,* **23,** 1, 19-35.
- Gómez-Villegas, M.A. and González-Pérez, B. (2011) A Bayesian analysis for the homogeneity testing problem using contaminated priors, *Communications in Statistics-Theory and Methods,* **40,** 6, 1049-1062.
- Ausín, M.C., Gómez-Villegas, M.A., González-Pérez, B., Rodríguez-Bernal, M., Salazar, I. and Sanz, L. (2011) Bayesian analysis of multiple hyphothesis testing with applications to Microarray experiments, *Communications in Statistics-Theory and Methods,* **40,** 13, 2276-2291.

- Gómez-Villegas, M.A.& Gómez, E. (1992) Bayes factors in testing precise hyphoteses, *Comm. Statist.-Theory Methods*, **21,** 1707-1715.

$$H_o : \theta = \theta_0 \quad \text{frente a} \quad H_1 : \theta \neq \theta_0$$

Con $P^*(A) = \pi_0 I_A(\theta_0) + (1-\pi_0)\int_A \pi(\theta)d\theta \qquad A \in \mathrm{B}_\Theta$

Es un caso límite de

$$H_{0\varepsilon} : \theta \in I_\varepsilon(\theta_0) = \{\theta \mid \; |\theta - \theta_0| \leq \varepsilon\} \text{ frente a } H_{1\varepsilon} : \theta \notin I_\varepsilon(\theta_0)$$

Con la distribución inicial $\pi(\theta)$

En el sentido que $\lim_{\varepsilon \to \infty} B_\varepsilon(x; H_{0\varepsilon}, H_{1\varepsilon}) = B(x; H_0, H_1)$

Si además $\pi_0 = \int_{I_\varepsilon(\theta_0)} \pi(\theta)d\theta$

Entonces $\lim_{\varepsilon \to \infty} P(H_{0\varepsilon} \mid x) = P\{H_0 \mid x\}$

Hipótesis puntual

Our procedure

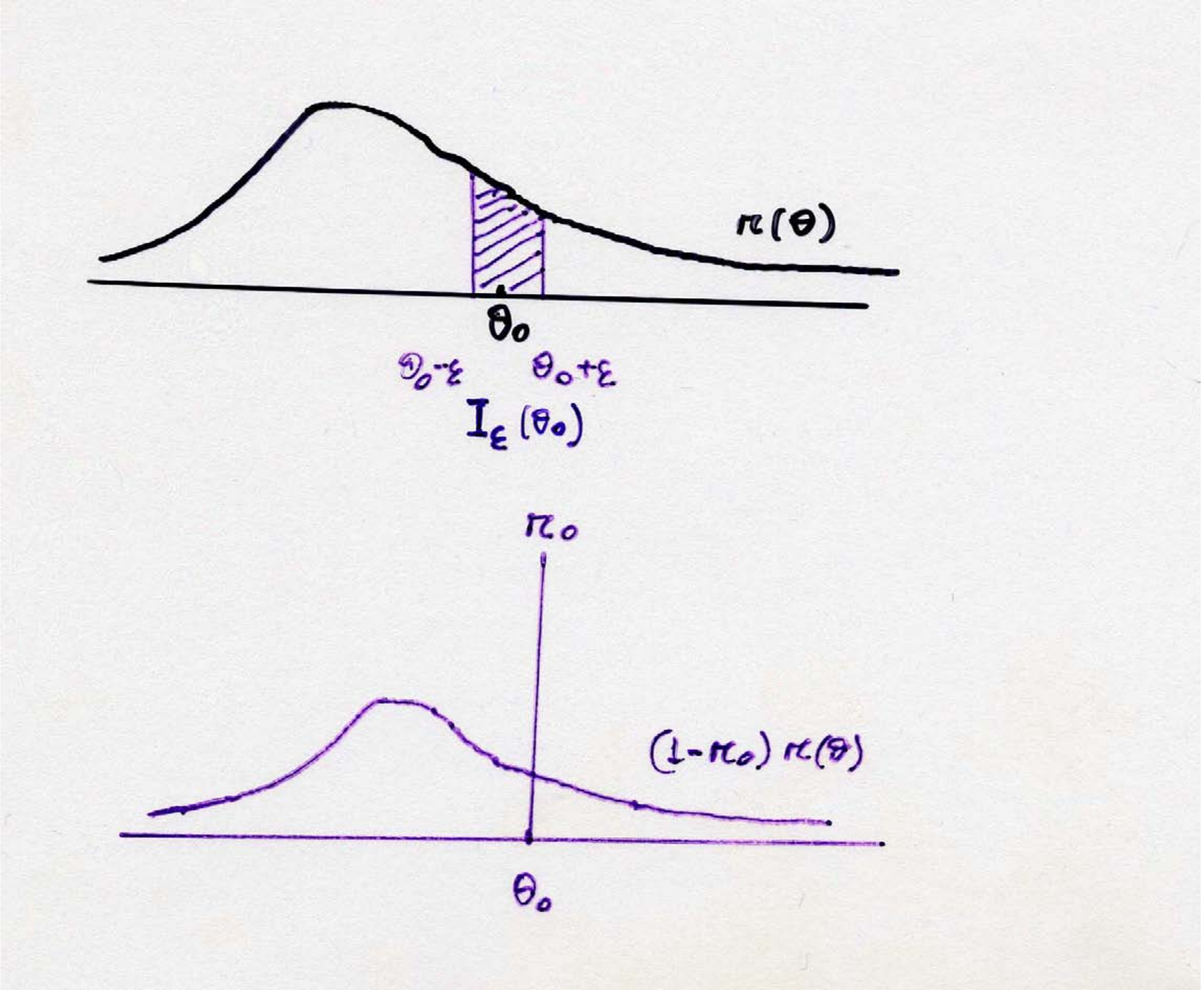
$$H_o : \theta = \theta_0 \quad \text{frente a} \quad H_1 : \theta \neq \theta_0$$

Cuando se emplea la distribución inicial

$$P^*(A) = \pi_0 I_A(\theta_0) + (1-\pi_0)\int_A \pi(\theta)d\theta \quad \text{con} \quad A \in \mathrm{B}_\Theta$$

Y se construye

$$\pi_0 = \int_{I_\varepsilon(\theta_0)} \pi(\theta)d\theta$$



$\pi(\theta)$
θ_0
$\theta_0-\varepsilon$
$\theta_0+\varepsilon$
$I_\varepsilon(\theta_0)$
π_0
$(1-\pi_0)\,\pi(\theta)$
θ_0

Theoretical justification

- La discrepancia de Kullback-Leibler

$$D(\pi^* \mid \pi) = \int_{\Theta} \pi(\theta) \ln \frac{\pi(\theta)}{\pi^*(\theta)} d\theta$$

- La discrepancia tiende a cero cuando ε tiende a 0
- Con π_0 = 0.5 la discrepancia vale 0.693
- En este caso no parece estar justificada una discrepancia simétrica

- Gómez-Villegas, M.A.and Sanz,L. (1998) Reconciling Bayesian and frequentist evidence in the point null testing problem.*Test,* **7**, 207-216.

<u>Teorema</u>. Para contrastar $H_0 : \theta = \theta_0$ frente a $H_1 : \theta \neq \theta_0$
con $f(x \mid \theta)$ continua en θ_0 si mediante una densidad

$\pi(\theta)$ se construye $\pi_0 = \int_{I_\varepsilon(\theta_0)} \pi(\theta) d\theta$ se tiene

$$\inf_{\pi \in \Gamma_{US}} P\{H_0 \mid x\} = \left[1 + \frac{1}{2\varepsilon} \int_{-\infty}^{+\infty} \frac{f(x \mid \theta)}{f(x \mid \theta_0)} d\theta \right]^{-1}$$

Existe un valor de ε tal que dicho ínfimo es igual al p-valor.
Con Γ_{US} la clase de las unimodales y simétricas alrededor de θ_0
Y no crecientes en $\mid \theta - \theta_0 \mid$

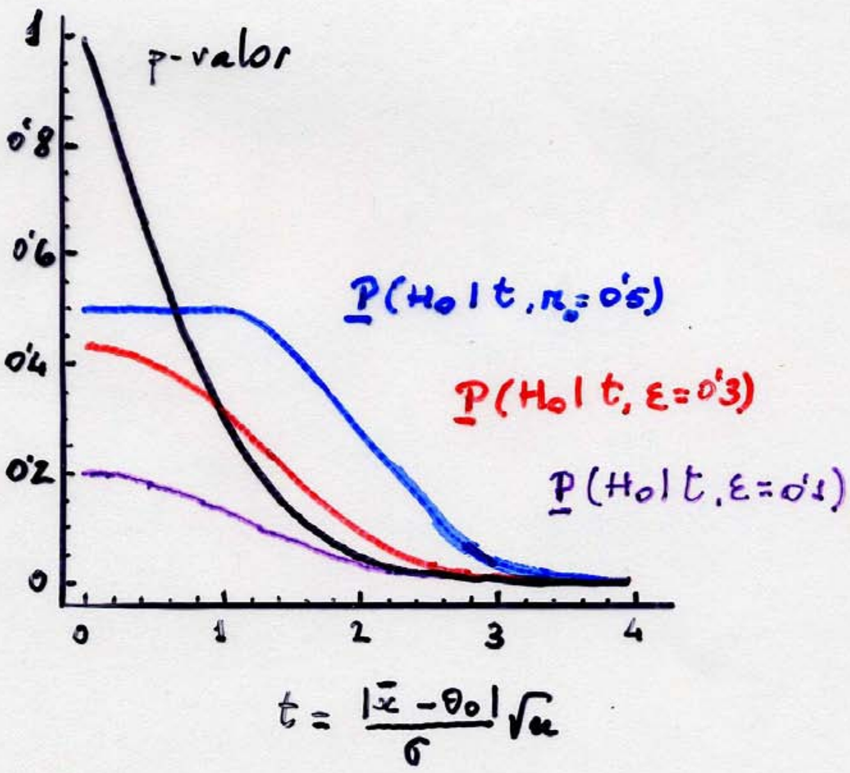


Figura: m.a.s. $n = 10$, $X \sim N(\theta, \sigma)$, $\sigma = 1$, $\theta_0 = 0$

Valores de épsilon que igualan el p-valor y el ínfimo de la probabilidad final para la clase de las distribuciones iniciales unimodales y simétricas

t	p-valor $= \underline{P}\left(H_0 \mid x\right)$	ε	$\underline{P}\left(H_{0\varepsilon} \mid x\right)$	$\underline{P}(H_0 \mid x, \pi_0 = 0.5)$
1.645	0.1	0.170	0.1198	0.39
1.96	0.05	0.142	0.0575	0.29
2.576	0.01	0.111	0.0121	0.109
3.291	0.001	0.089	0.0011	0.018

- Gómez-Villegas, M.A. & Sanz, L. (2000) épsilon-contaminated priors in testing point null hypothesis: a procedure to determine the prior probability. *Statistics & Probability Letters,* **47,** 53-60.

Se utiliza la clase de las distribuciones ε contaminadas

$$\Gamma_{\varepsilon c} = \left\{ \pi(\theta) \mid \text{ con } \pi = (1-\varepsilon)\pi_0 + \varepsilon q, \; q \in Q \right\}$$

Y se construye

$$\pi^*(\theta) = pI_{\{\theta_0\}}(\theta) + (1-p)\pi(\theta)I_{\{\theta \neq \theta_0\}}(\theta)$$

$$p = \int_{I_b(\theta_0)} \pi(\theta)d\theta$$

Se puede conseguir acuerdo entre ambas aproximaciones para Q la clase de todas las distribuciones.

- Gómez-Villegas, M. A., Maín, P. and Sanz, L. (2002) A suitable Bayesian approach in testing point null hyphotesis: some examples revisited. *Commun. Statist.-Theory Meth.,* **31**, 2, 201-217.

¿Qué ocurre si se comparan directamente $P\left\{\theta = \theta_0 \mid \vec{x}\right\}$ y el p-valor?

-Paradoja de Lindley

-Inicial Cauchy y el modelo Cauchy

-Problema de Darwin-Fisher

Es posible obtener un valor de ε tal que ambos valores sean próximos

Valores de épsilon que igualan la probabilidad final y el p-valor, con

X $\sim$ Cauchy(θ,1) y $\theta \sim$ Cauchy(0,2)

x	ε	$P\{H_0 \mid x\} = p(x)$	$P\{H_{0\varepsilon} \mid x\}$
1.0	4.84	0.500	0.998
1.5	3.67	0.374	0.966
3.0	2.76	0.205	0.638
6.3	1.64	0.100	0.185
12.7	0.87	0.050	0.092

Valores de épsilon que igualan el p-valor a la probabilidad final para el problema de Darwin-Fisher

s_0	ν_0	ν_1	ε	$p(t) \approx P\{H_0 \mid \varepsilon, \vec{x} = 2.617\}$	$P\{H_{0\varepsilon} \mid t\}$	$P\{H_0 \mid \pi_0 = 0.5\}$
1	7	20	0.121	0.05	0.0481	0.3399
2.5	7	20	0.379	0.05	0.0489	0.2869
10	7	20	0.680	0.05	0.0579	0.4889
125	7	20	0.747	0.05	0.0625	0.9193

- Gómez-Villegas, M.A., Maín, P., Sanz, L. and Navarro, H. (2004) Asymptotic relationship between posterior probabilities and p-values using the hazard rate. *Statististics & Probability Letters,* **66**, 59-66.

- Portela, J. and Gómez-Villegas, M.A. (2004) Implementation of a robust Bayesian method. *Journal of Statistical Computation and Simulation,* **74**, 4, 235-248.

- Gómez-Villegas, M.A., Portela, J. and Sanz, L. (2008) A Bayesian test for the mean of the Power Exponential distribution. *Communications in Statistics-Theory and Methods,* **37**, 18, 2865-2876.

Gómez-Villegas, M.A. and González Pérez, B. (2006) Bayesian analysis of contingency tables, *Commun. Stattist. Theor. Meth.*, **34,** 1743-1754,

	éxitos	fallos	
	a	b	n_1
	c	d	n_2
	m_1	m_2	N

- En una tabla de contingencia

  Se pretende contrastar

$$H_0 : p_1 = p_1^0, p_2 = 1 - p_1^0$$

frente a $H_1 : p_1 = q_1, p_2 = 1 - q_1$

- Desde el punto de vista frecuentista se emplea

RC= $\left\{ \Lambda = \sum {(O_i - E_i \mid H_i)^2} \big/ {E_i \mid H_i} \right\}$

p-valor=P{RC|Datos}

	éxitos	fallos	
	3	15	18
	7	5	12
	10	20	30

Para hacer las comparaciones se recurre a los datos de K.Pearson(1947)

p-valor= $P\{\Lambda \geq \lambda\} = P\{\chi_1^2 \geq \lambda = 8.167\}$
con lo que el p-valor es 0.0043
En este caso, para los valores
usuales de $\alpha = 0.01, 0.05$ y 0.1 se rechaza la hipótesis nula.

- Desde el punto de vista bayesiano se emplea distribución inicial

$$\pi^*(p_1, p_2) = \pi_0 I_{H_0}(p_1, p_2) + (1-\pi_0)\pi(p_1, p_2) I_{H_1}(p_1, p_2)$$

con

$$\pi(p_1, p_2) = Beta(p_1 \mid \alpha, \beta) Beta(p_2 \mid \gamma, \delta)$$

$$\pi_0 = \int_{B((p_1^0, p_2^0), \varepsilon)} \pi(p_1, p_2) dp_1 dp_2$$

y B la bola de centro (p_1^0, p_2^0) y radio ε = 0.399

probabilidad final = $\left[1 + \frac{(1-\pi_0)}{\pi_0}\eta\right]^{-1}$

Por lo que la probabilidad final es 0.1294 y se rechaza la hipótesis nula.

Gómez-Villegas,M.A. and González,B. (2006) A condition to obtain the same decision in the homogeneity testing problem from the frequentist and Bayesian point of view, *Communications in Statistics-Theory and Methods*, **35**, 2211-2222.

- En general es posible hacer que la probabilidad final y el p-valor estén próximos

- Basta calcular

$$l_1 = \max_{p > p^*} \left\{ \eta(\eta+1)^{-1} \right\} \qquad l_2 = \min_{p \leq p^*} \left\{ \eta(\eta+1)^{-1} \right\}$$

y que se tenga

$$l_1 < \pi_0 < l_2$$

- En otro caso no es posible lograr el acuerdo.

Last comments

- Our procedure allows the matching of the $\boldsymbol{p(\vec{x})}$ –value and the $\underline{P}\left\{H_0 \mid \vec{x}\right\}$
- The behaviour of the punctual hypothesis is coherent with the one-sided hypothesis
- The difference between the frequentist approach and Bayesian approach is originated by making use of

$$\pi_0 = 0.5$$

- Gómez-Villegas, M.A., Maín, P. & Sanz, L.(2010) A Bayesian analysis for the multivariate point null testing, *Statistics*, **43**, 4, 379-391
- Ausín, M.C., Gómez-Villegas, M.A., González-Pérez, B., Rodríguez-Bernal,M.T., Salazar, I. & Sanz, L.(2011) Bayesian analysis of multiple hypothesis with applications to Microarray experiments, *Communications in Statistics-Theory and Methods,* **40**, 13, 2276-2291
- Generalization to more general problems : Behrens-Fisher, Hardy-Weinberg

Here, everybody is Bayesian, but…

- Gómez Villegas, M.A. (2005, 2011) *Inferencia Estadística*, Madrid: Díaz de Santos

Thank you¡