

Comunicaciones del CIMAT

INFERENCIA SOBRE EL COCIENTE
DE DOS MEDIAS NORMALES

Francisco Javier Rubio Alvarez

Comunicación del CIMAT No I-08-02/17-01-2008
(PICIMAT)



CIMAT

Inferencia Sobre el Cociente de Dos Medias Normales

Francisco Javier Rubio Alvarez
CIMAT, Jalisco S/N, Colonia Valenciana
36240, Guanajuato, Gto, México.
fjavier@cimat.mx

Resumen

Se presentan de manera resumida los trabajos de Mendoza & Gutiérrez-Peña [1] y de Díaz-Francés & Sprott [3] acerca de la inferencia bayesiana y frecuentista sobre el cociente de dos medias normales, además se muestra un error numérico en el trabajo de M & G-P [1] así como las consecuencias de este error.

1 Introducción

El problema de hacer inferencia sobre la media del cociente de dos poblaciones normales ha sido discutido ampliamente tanto en el contexto bayesiano como en el frecuentista. Los objetivos del presente trabajo son:

- i) Presentar de manera resumida los resultados de un artículo de Mendoza & Gutiérrez-Peña [1] sobre la inferencia bayesiana del cociente de medias de dos poblaciones normales para el modelo de Cox y para el caso general.
- ii) Presentar un error numérico y sus consecuencias en el artículo de M & G-P [1].
- iii) Presentar la crítica hecha por Díaz-Francés & Sprott [3] a este artículo y la opción frecuentista para analizar el problema.

2 Contexto Bayesiano

Primero analizaremos el trabajo hecho por M & G-P [1], en este trabajo hacen una crítica a la opción frecuentista, que se discutirá más adelante, esta crítica se hace principalmente a la aplicación del pivotal de Fieller (esto es, la construcción de la verosimilitud a partir de una cantidad pivotal) ya que en ciertos casos puede llevar a intervalos de verosimilitud de longitud infinita. El modelo de la sección 2.1 está basado en el caso presentado por Cox [2]. En la sección 2.2 se analiza el caso general.

2.1 Modelo de Cox

Sean $X = (X_1, \dots, X_n)$ y $Y = (Y_1, \dots, Y_m)$ dos muestras independientes tales que:

$$\begin{aligned} X_i &\sim N(x|\mu, \sigma_1^2), \\ Y_i &\sim N(y|\eta, \sigma_2^2), \end{aligned}$$

con $\sigma_1^2 = \varrho^2(c + \mu)^k$, y $\sigma_2^2 = \varrho^2(c + \eta)^k$, donde c y k son constantes conocidas, c es elegida de tal manera que las varianzas sean positivas. En el artículo de Cox [2] se propone un pivotal (basado en la fórmula de Fieller) para obtener intervalos de confianza para $(c + \mu)/(c + \eta)$.

Entonces, si $c = 0$, esta inferencia corresponde al cociente de la medias. Entonces supondremos $c = 0$ y consideraremos el caso $k = 2$ y medias positivas, con esto estamos suponiendo que ambas poblaciones tienen el mismo coeficiente de variación (las poblaciones tienen la misma dispersión).

El parámetro de interés es $\varphi = \mu/\eta$, la función de verosimilitud para los parámetros (φ, η, ϱ) , está dada por:

$$\begin{aligned} L(\varphi, \eta, \varrho) &\propto \varrho^{-(n+m)} \varphi^{-n} \eta^{-(n+m)} \times \\ &\exp \left\{ -\frac{1}{2\varrho^2 \varphi^2 \eta^2} \left(\sum_{i=1}^n (x_i - \varphi\eta)^2 + \varphi^2 \sum_{j=1}^m (y_j - \eta)^2 \right) \right\}. \end{aligned} \quad (1)$$

La inferencia bayesiana para φ está basada en la distribución marginal posterior. Entonces debemos asignar una distribución *a priori* conjunta para el vector de parámetros (φ, η, ϱ) . En la práctica se tendría un problema si queremos que esta distribución *a priori* provea información genuina. En este caso, como no se tiene un contexto M & G-P [1] dan una *a priori* de referencia. Calculando esta *a priori* resulta:

$$\pi(\varphi, \eta, \varrho) \propto \varphi^{-1} \eta^{-1} \varrho^{-1},$$

donde cada parámetro es positivo por hipótesis. Desafortunadamente esta *a priori* lleva a una posterior impropia para cualesquiera tamaños de muestra. Sin embargo este resultado sugiere una *a priori* de pseudoreferencia de la forma:

$$\pi(\varphi, \eta, \varrho) \propto \varphi^{-a} \eta^{-b} \varrho^{-c}.$$

Es suficiente tomar $a = 1$, $b = 1/2$, $c = 1$, para llegar a una posterior propia. Entonces la distribución posterior correspondiente a esta elección es la siguiente:

$$\begin{aligned} \pi(\varphi, \eta, \varrho | \mathbf{x}, \mathbf{y}) &\propto \varrho^{-(n+m+1)} \varphi^{-(n+1)} \eta^{-(n+m+1/2)} \times \\ &\exp \left\{ -\frac{1}{2\varrho^2} \left(\frac{n}{\varphi^2 \eta^2} \{s_x^2 + (\varphi\eta - \bar{x})^2\} + \frac{m}{\eta^2} \{s_y^2 + (\eta - \bar{y})^2\} \right) \right\}. \end{aligned} \quad (2)$$

Integrando con respecto de ϱ^2 y relacionando el integrando con la densidad gama inversa (haciendo el cambio de variable $U = \varrho^2$) podemos eliminar este parámetro y obtener la siguiente marginal:

$$\pi(\varphi, \eta | \mathbf{x}, \mathbf{y}) \propto \varphi^{-(n+1)} \eta^{-(n+m+1/2)} \times \left[\frac{n}{\varphi^2 \eta^2} \{s_x^2 + (\varphi \eta - \bar{x})^2\} + \frac{m}{\eta^2} \{s_y^2 + (\eta - \bar{y})^2\} \right]^{-(n+m)/2}. \quad (3)$$

El parámetro η no puede ser eliminado de la misma manera, así que para obtener la distribución posterior marginal $\pi(\varphi | \mathbf{x}, \mathbf{y})$, es necesario hacer una aproximación numérica integrando para diferentes valores del parámetro φ .

M & G-P [1] presentan la posterior de $\delta = \sigma_1/\mu - \sigma_2/\eta$, para la cual utilizan esta misma *a priori* de referencia, se reproduce a continuación en la Figura 1.

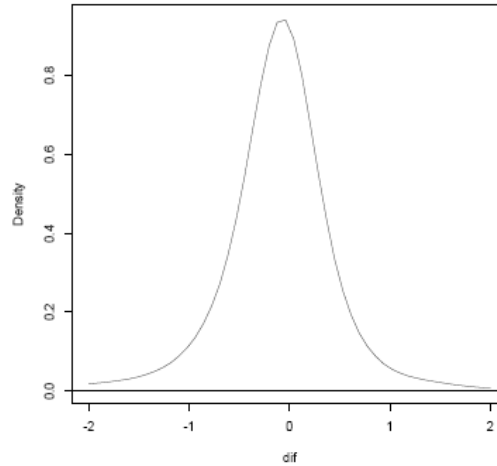


Figura 1: Posterior de M & G-P para δ .

Por lo tanto no hay evidencia en contra de que $\delta = 0$, luego, no hay evidencia en contra del modelo de Cox.

Entonces, encontramos la constante normalizadora e integrando con el método de Gauss-Kronrod (el cual consiste en el método de cuadratura de Gauss y utiliza los puntos de Kronrod) para 500 puntos se obtiene la siguiente aproximación a la densidad posterior de φ en Mathematica 5.0:

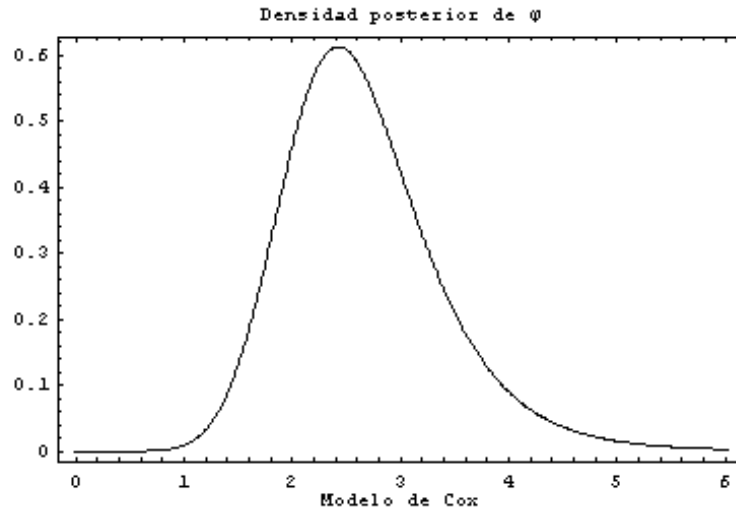


Figura 2 : Densidad posterior de φ , Modelo de Cox.

A continuación se reproduce la densidad posterior de φ mostrada en el artículo de M & G-P [1] para el modelo de Cox:

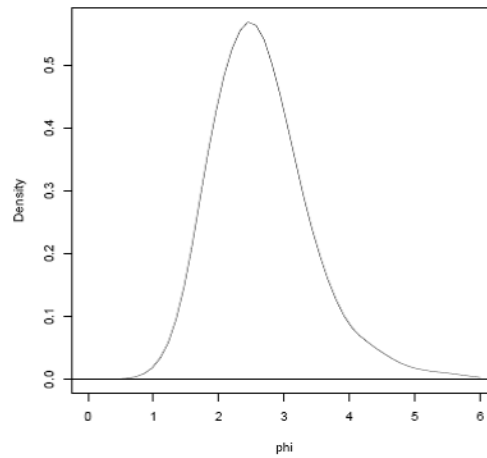


Figura 3 : Posterior de φ de M & G-P, Modelo de Cox.

En el artículo de M & G-P [1] utilizan el algoritmo de Metropolis-Hastings y 2000 iteraciones, pero debido a que la integral no es muy complicada, puede

ser encontrada de forma más sencilla con métodos de cuadratura. Nótese de la Figura 3 que la cola derecha sugiere un error numérico en su método, esto es debido al bajo tamaño de muestra que utilizaron.

Aquí el intervalo HPD de 95% de probabilidad reportado por M & G-P [1] es [1.26, 4.16], lo cual coincide aproximadamente con el HPD encontrado mediante integración numérica con cuadratura, [1.33, 4.17].

2.2 Modelo General.

En el artículo de Cox se propone un modelo para un valor arbitrario de k pero sólo trata a detalle el caso $k = 2$. M & G-P [1] proponen que k sea un parámetro más en el modelo. Si el parámetro k es desconocido la validez de la relación media-varianza está abierta a la crítica. Como alternativa, el modelo de Cox puede ser visto como un caso particular de la situación general en el que las varianzas no son iguales, cuya solución es más robusta en el sentido de que no tiene hipótesis estructurales.

Entonces se tiene que las varianzas σ_1^2 , σ_2^2 son desconocidas y no se supone relación alguna con las medias. Para este modelo la verosimilitud es:

$$L(\varphi, \eta, \sigma_1, \sigma_2) \propto \sigma_1^{-n} \sigma_2^{-m} \exp \left\{ -\frac{1}{2} \left(\frac{1}{\sigma_1^2} \sum_{i=1}^n (x_i - \varphi\eta)^2 + \frac{1}{\sigma_2^2} \sum_{j=1}^m (y_j - \eta)^2 \right) \right\}. \quad (4)$$

La correspondiente *a priori* de referencia es:

$$\pi(\varphi, \eta, \sigma_1, \sigma_2) \propto \varphi^{-1/2} \sigma_1^{-1} \sigma_2^{-1},$$

donde todos los parámetros son positivos. Esta *a priori* produce la siguiente posterior:

$$\begin{aligned} \pi(\varphi, \eta, \sigma_1, \sigma_2 | \mathbf{x}, \mathbf{y}) &\propto \varphi^{-1/2} \sigma_1^{-(n+1)} \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{j=1}^n (x_j - \varphi\eta)^2 \right\} \\ &\quad \times \sigma_2^{-(m+1)} \exp \left\{ -\frac{1}{2\sigma_2^2} \sum_{j=1}^m (y_j - \eta)^2 \right\}. \end{aligned} \quad (5)$$

Aquí sí se tiene una posterior propia, podemos integrar directamente sobre σ_1 y σ_2 y obtener la posterior conjunta de (φ, η) . Para obtener la posterior de φ tendrá que hacerse nuevamente aproximación numérica. Para hacer la primera integral relacionamos esta expresión con la densidad gama inversa y hacemos los cambios de variable $U_i = \sigma_i^2$, con esto obtenemos la siguiente densidad conjunta *a posterior*:

$$\pi(\varphi, \eta | \mathbf{x}, \mathbf{y}) \propto \varphi^{-1/2} \left[\frac{1}{2} \sum_{j=1}^n (x_j - \varphi \eta)^2 \right]^{-n/2} \left[\frac{1}{2} \sum_{j=1}^n (y_j - \eta)^2 \right]^{-m/2}. \quad (6)$$

Normalizando y utilizando el método de integración de Gauss-Kronrod con 800 puntos se obtiene la aproximación para la densidad posterior de φ que se muestra en la Figura 4.

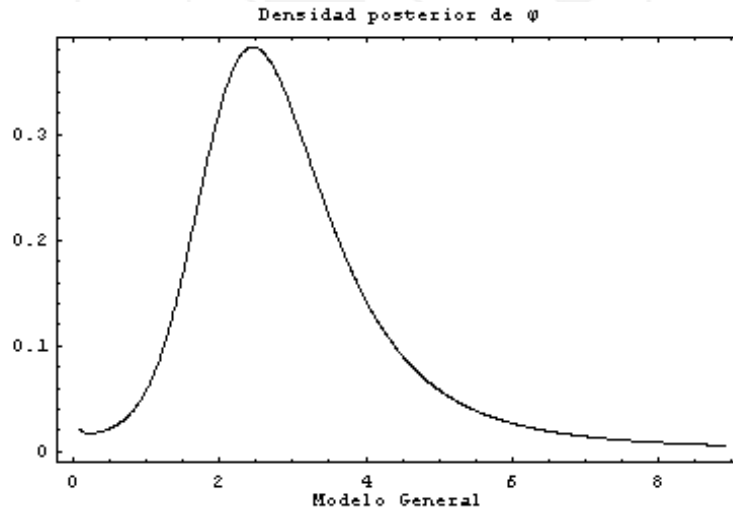


Figura 4 : Densidad Posterior de φ , Modelo General.

En la Figura 5 se reproduce la densidad posterior de φ del modelo general reportada por M & G-P [1].

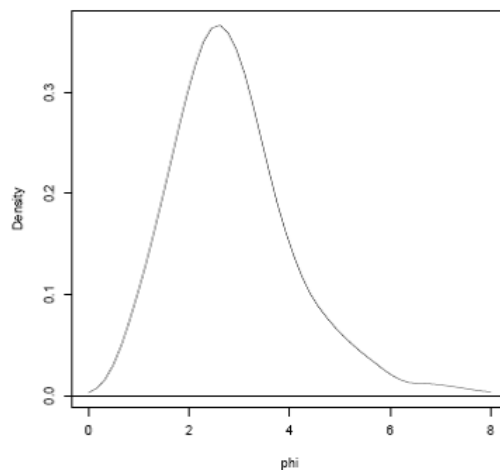


Figura 5 : Densidad posterior de φ , M & G-P [1], Modelo General.

M & G-P [1] reportan el siguiente intervalo como HPD del 95% $[0.63, 5.29]$, cuya longitud es 4.66.

En primer lugar nótese la diferencia en ambas colas de las densidades, nuevamente, esto es debido a la muestra tan pequeña que utilizaron, además notemos la "irregularidad" de la gráfica lo cual nos sugiere la presencia de errores numéricos.

Utilizando el método de Gauss-Kronrod aproximamos la integral (en Mathematica 5.0) de la marginal en el intervalo $[0.63, 5.29]$, reportado por M & G-P [1] como de 95% de probabilidad, y resulta ser de 0.887837, lo cual indica una falta total de aproximación del método usado por M & G-P [1].

De hecho haciendo la integral por el método de MonteCarlo de Mathematica pero utilizando 100 000 iteraciones se obtiene 0.883882 lo cual concuerda con el método de Gauss-Kronrod.

Es fácil que en alguna evaluación, el intervalo $[0.63, 5.29]$ resulte del 95% con 2000 iteraciones utilizando el método de MonteCarlo de Mathematica 5.0. Para ver la variación de los resultados de la probabilidad de este intervalo utilizando el método de MonteCarlo de Mathematica 5.0 con 2000 iteraciones, se realizó esta misma integral 100 000 veces, el histograma de los resultados obtenidos se muestra en la Figura 6.

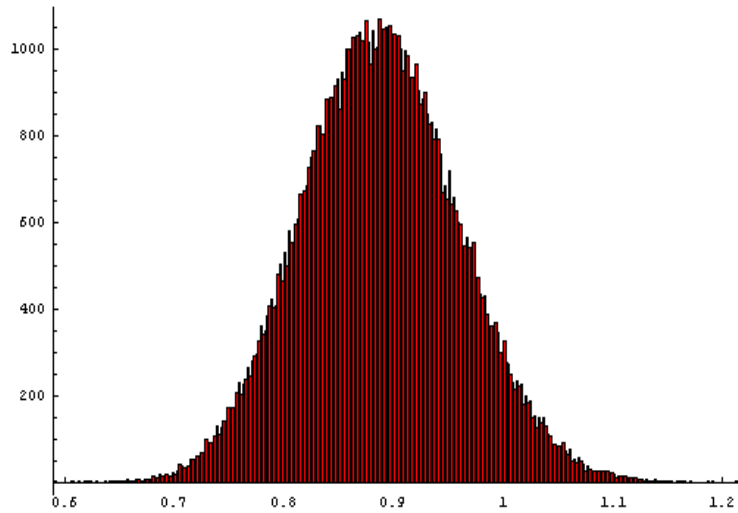


Figura 6.

Como podemos ver la variación de los resultados es muy grande como para confiar en el resultado obtenido con 2000 iteraciones, además esto explica el resultado reportado por M & G-P [1]. La razón por la cual algunas veces la integral da mayor que 1, es debido a que el método de Montecarlo puede dar resultados mayores ó menores al Gauss-Kronrod, y la constante normalizadora fué calculada con el método de Gauss-Kronrod.

Ahora, realizamos la integral 1000 veces pero ahora utilizando 100 000 iteraciones para el método de MonteCarlo de Mathematica 5.0, el histograma de

los resultados obtenidos en este caso se muestra en la Figura 7.

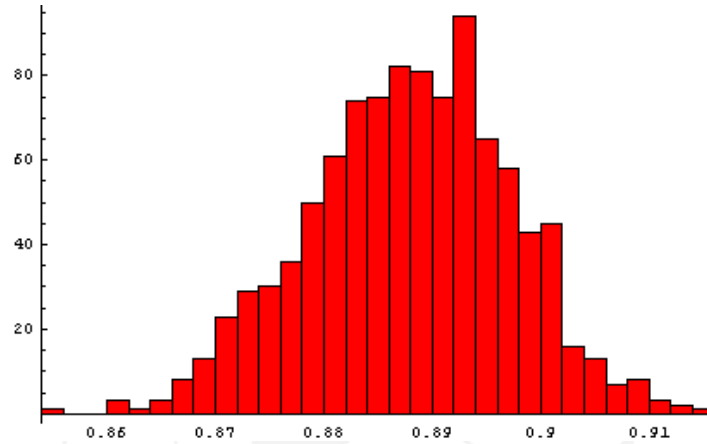


Figura 7.

Podemos ver que la variación de los resultados es mucho menor al aumentar la muestra lo cual confirma la aseveración de que 2000 es un tamaño de muestra pequeño en este problema. Cabe mencionar que dado que el método de integración de Gauss-Kronrod es un método de cuadratura, este no tiene variación entre una evaluación y otra ya que no se basa en cantidades aleatorias.

El intervalo HPD del 95% calculado con Gauss-Kronrod y Newton-Raphson es $[0.191, 7.740]$, cuya longitud es 7.55, integrando sobre este intervalo con Gauss-Kronrod y con MonteCarlo utilizando 100 000 iteraciones se obtiene 0.95 como era de esperarse, mientras que con un tamaño de 2000 se tienen altas variaciones.

Textualmente una conclusión del artículo de M & G-P [1] es "for the fine gravel data set in the example, the HPD interval for the ratio of the means are more precise under Cox's model (with $k=2$) than they are under the general model. More interestingly, the Bayesian interval for the general model is narrower than the corresponding confidence interval obtained by Cox for his model."

El intervalo que da Cox [2] en su modelo es $[1.27, 6.68]$, cuya longitud es 5.41, bajo el modelo general corresponde a un intervalo de 0.891158 de probabilidad. Con esto, se está cometiendo un error al hacer la conclusión mencionada en el párrafo anterior sobre la longitud de los intervalos, **nos parece que esta conclusión es una deducción realizada a partir del error numérico anteriormente mencionado.**

3 Contexto Frecuentista

En el contexto frecuentista lo mejor es utilizar la verosimilitud completa, los modelos de Cox construyen la verosimilitud basada en el pivotal de Fieller, como

veremos a continuación, esta verosimilitud es sólo una parte de la verosimilitud completa por lo tanto no tiene toda la información de los parámetros involucrados. Estos resultados están basados en un artículo de Díaz-Francés & Sprott [3].

Sean $X = (X_1, \dots, X_n)$ y $Y = (Y_1, \dots, Y_m)$ dos muestras independientes de distribuciones normales tales que:

$$\begin{aligned} x_i &\sim N(\beta\mu, \rho^2\sigma^2) & i = 1, \dots, n, \\ y_i &\sim N(\mu, \sigma^2) & i = 1, \dots, m. \end{aligned} \quad (7)$$

El pivotal de Fieller usado para hacer inferencia sobre β como función de ρ es:

$$t^2 = \frac{nm(\bar{x} - \beta\bar{y})^2}{(m\rho^2 + n\beta^2)\hat{\sigma}^2},$$

$$\text{donde } (n+m-2)\hat{\sigma}^2 = S_x/\rho^2 + S_y, \quad S_x = \sum_{j=1}^n (x_j - \bar{x})^2, \quad S_y = \sum_{j=1}^m (y_j - \bar{y})^2.$$

El caso considerado por M & G-P [1] es en particular cuando $\rho = \beta$.

3.1 La función de Verosimilitud

Bajo el modelo (7) la función de verosimilitud es:

$$\begin{aligned} L(\beta, \mu, \rho, \sigma; X, Y) &\propto \frac{1}{\rho^n \sigma^{n+m}} \exp\left(-\frac{Q}{2\sigma^2}\right), \\ Q &= \left[\frac{S_x}{\rho^2} + S_y + \frac{n(\bar{x} - \beta\bar{y})^2}{\rho^2} + m(\bar{y} - \mu)^2 \right]. \end{aligned}$$

El estimador máximo verosímil para β y ρ especificadas es:

$$\hat{\mu}(\beta, \rho) = \frac{n\beta\bar{x} + m\rho^2\bar{y}}{n\beta^2 + m\rho^2},$$

entonces la verosimilitud perfil de β, σ, ρ es:

$$\begin{aligned} L_p(\beta, \rho, \sigma; X, Y) &\propto \frac{1}{\rho^n \sigma^{n+m}} \exp\left(-\frac{C}{2\sigma^2}\right), \\ C &= \left[\frac{S_x}{\rho^2} + S_y + \frac{nm(\bar{x} - \beta\bar{y})^2}{n\beta^2 + m\rho^2} \right]. \end{aligned}$$

Análogamente,

$$\hat{\sigma}^2(\beta, \rho) = \frac{C}{m+n}.$$

La perfil de β, ρ es:

$$\begin{aligned} L_p(\beta, \rho; X, Y) &\propto \frac{1}{\rho^n} C^{-\frac{m+n}{2}}, \\ &= \rho^m \left[S_x + \rho^2 S_y + \frac{\rho^2 nm(\bar{x} - \beta\bar{y})^2}{n\beta^2 + m\rho^2} \right]^{-\frac{m+n}{2}}. \end{aligned}$$

Esta verosimilitud puede ser factorizada como:

$$L_p(\beta, \rho) = L_F(\rho) L_{Fieller}(\beta|\rho),$$

donde

$$\begin{aligned} L_F(\rho) &\propto \rho^m (S_x + \rho^2 S_y)^{-\frac{m+n}{2}}, \\ L_{Fieller}(\beta|\rho) &\propto \left[1 + \frac{\rho^2 nm(\bar{x} - \beta\bar{y})^2}{(S_x + \rho^2 S_y)(n\beta^2 + m\rho^2)} \right]^{-\frac{m+n}{2}}. \end{aligned}$$

En el caso analizado por M & G-P [1] $\rho = \beta$, entonces:

$$\begin{aligned} L(\beta) &\propto \beta^m \left[S_x + \beta^2 S_y + \frac{nm(\bar{x} - \beta\bar{y})^2}{n + m} \right]^{-\frac{m+n}{2}}, \\ L_F(\beta) &\propto \beta^m (S_x + \beta^2 S_y)^{-\frac{m+n}{2}}, \\ L_{Fieller}(\beta) &\propto \left[1 + \frac{nm(\bar{x} - \beta\bar{y})^2}{(S_x + \beta^2 S_y)(n + m)} \right]^{-\frac{m+n}{2}}. \end{aligned}$$

Notemos que:

$$\begin{aligned} \lim_{\beta \rightarrow \pm\infty} L_F(\beta) &= 0, \\ \lim_{\beta \rightarrow \pm\infty} L_{Fieller}(\beta) &> 0. \end{aligned}$$

En la Figura 8 se presentan las verosimilitudes perfil $L(\beta)$, $L_F(\beta)$ y $L_{Fieller}(\beta)$

obtenidas con los datos presentados por M & G-P [1].

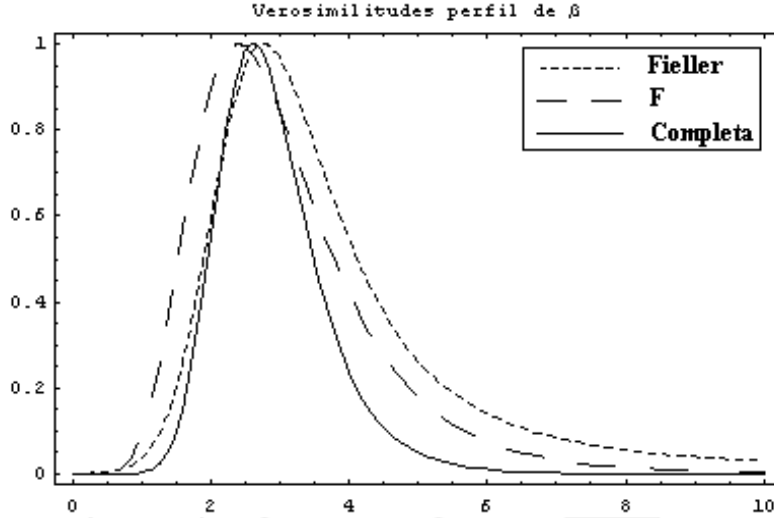


Figura 8.

El intervalo de 0.1 de verosimilitud calculado con la verosimilitud completa es $[1.51, 4.54]$, el cual es comparable con el $[1.26, 4.16]$ encontrado por M & G-P [1].

En el artículo de M & G-P [1] se critican los intervalos obtenidos a partir del pivotal de Fieller. La crítica es válida, ya que al usar sólo la parte de la verosimilitud $L_{Fieller}(\beta)$, se está dejando de lado la información de β contenida en $L_F(\beta)$ la cual es igual de informativa sobre éste parámetro, lo cuestionable sobre esto es que comparan el método bayesiano con los resultados obtenidos utilizando sólo $L_{Fieller}(\beta)$, es decir con una parte (que de entrada se sabe que dará resultados erróneos) del método frecuentista.

4 Conclusiones

A partir del error numérico presente en el trabajo de M & G-P [1] se llega a la conclusión errónea de que el modelo general bayesiano produce intervalos más angostos que el modelo particular de Cox [2]. La importancia de este error no radica sólo en los valores obtenidos para el intervalo de probabilidad, sino en las conclusiones derivadas por M & G-P [1] a partir de este error.

Díaz-Francés & Sprott [3] muestran que en el caso presentado por M & G-P [1], la opción bayesiana y la frecuentista producen resultados similares en cuanto a los intervalos de probabilidad y de confianza. Cabe resaltar que las densidades *a priori* utilizadas por M & G-P [1] son *a priori* de referencia y por lo tanto no hay un aporte de información genuina.

En el trabajo presentado por Díaz-Francés & Sprott [3] se muestra que al utilizar sólo una parte residual de la función de verosimilitud para hacer inferencia sobre el parámetro de interés se puede llegar a inferencias malas tales como intervalos de verosimilitud de longitud infinita.

5 Agradecimientos

El presente trabajo fue realizado como trabajo final del curso de Estadística Bayesiana impartido por el Dr. J. Andrés Christen en el CIMAT en el periodo agosto-diciembre de 2007.

Agradezco al Dr. J. Andrés Christen por su orientación y valiosos comentarios para la realización de este trabajo.

6 Referencias

- [1] "Bayesian Inference for the ratio of the means of two normal populations with unequal variances", M. Mendoza & E. Gutiérrez-Peña.
- [2] "Interval Estimates for the Ratio of the Means of Two Normal Populations with Variances Related to the Means", C. Philip Cox.
- [3] "Inferences for the Ratio of Two Normal Means with Unspecified Variances", Eloísa Díaz-Francés & David A. Sprott.