El Conflicto en Inferencia Estadística Verosimilitud

Kjetil Brinchmann Haslvorsen

Contrario a o que piensan muchas personas no iniciada, los fundamentos para la Inferencia Estadística no son resueltos para la satisfacción de todos, pero es una fuente de mucho conflicto . Lo que ahora se llama "Estadística Clásica", con raíces desde los años 1920, han proporcionado métodos muy usados como pruebas de significancia pura, prueba de hipótesis , intervalos de confianza entre otros. Esto sin duda ha sido un gran aporte práctico a la ciencia, pero su base lógica no es muy firme. En el mismo año (1934) como Newman & Pearson propuso la teoría de intervalos de confianza, Fisher lo atacaba fuertemente. ¿Porqué? no satisface los requerimientos de la teoría de la inducción científica, no produce una medida de evidencia en un cierto cuerpo de datos sobre alguna teoría científica, pero reemplaza en el concepto de evidencia por el de "la propiedad de confianza". Este último básicamente dice que no tenemos confianza en un cierto intervalo de confianza, pero tenemos sólo confianza en el método que lo ha producido. En hecho algunos estadísticos modernos muy conocidos como Lucien LeCam, argumenta que no necesariamente existe evidencia acerca de una hipótesis científica, basado en un experimento concreto"(Lucien LeCam, en su discusión en Berger & Wolpert (2).

Para la mayoría de los científicos en el campo aplicado, este argumento parece raro, por decir lo menos.

Existe alternativas a la estadística Clásica, y los más importantes son la Inferencia Bayesiana y la Inferencia de Verosimilitud Pura. Estos son paradigmas alternativos, aplicables para todo el campo de la estadística. Otras teorías, menos conocidas, tienen aparentemente menor aplicabilidad, y no las mencionaremos en este artículo. Para una discusión y referencias, mire por ejemplo Edwards (4) ó Hacking (6).

¿Porqué debemos preocuparnos de estos temas aquí, en lugar de sólo preocuparnos de los aspectos inmediatamente aplicables a la Estadística? Por múltiples razones, primero por su importancia e interés, segundo, por razones pedagógicas. Talvez la mejor manera de entender los a los métodos y principios estadísticos, es presentar diferentes teorías de una manera comparativa, con énfasis en la diferencia en base lógica e interpretación . este argumento es particularmente apropiado en este momento cuando la Carrera de estadística de la Universidad Mayor de San Andrés se encuentra en una fase de discusiones de pedagogía y metodología, por razón del Diplomado en Educación Superior.

La evidencia es que la mayoría de los usuarios de intervalos de confianza lo interpretan de una manera evidencial, que no es estrictamente permitida. Por ejemplo en Berger (1) Berger dice que ha tenido éxito en enseñar intervalos de confianza a estudiantes en cursos introductorios, sólo cuando paralelamente ha enseñado intervalos de confianza a estudiantes en cursos introductorios, cuando paralelamente ha enseñado intervalos bayesianos.

Entonces, existen muchas buenas razones para una enseñanza "agnóstica" de los principios de la Estadística. Este autor ha presentado en artículos anteriores aspectos de la teoría bayesiana en nuestro medio, por ejemplo (7), y en otros documentos. Ahora es tiempo para una pequeña presentación de otro punto de vista alternativo, sólo con aspectos de la verosimilitud .

2. PROBLEMAS CON LA TEORÍA CLÁSICA

Presentamos un solo ejemplo sencillo, tomando (2), mostrando el tipo de problemas que ocurren con la teoría clásica de intervalos de confianza. Básicamente, ocurren problemas porque las inferencias clásicas no son condicionales en los datos.
En los datos.

Sea X_1 y X_2 independientes, cada uno con la distribución.

$$P(X_i = \theta - 1 \mid \theta) = P(X_i = \theta + 1 \mid \theta) = \tfrac{1}{2}, \quad i = 1,2$$

Aquí $-\infty < \theta < \infty$ es un parámetro desconocido, para estimar en base a las observaciones. Es fácil mostrar que un conjunto de confianza con grado de confianza 75% esta dado por
donde el grado de confianza es 75% quiere decir que, en uso repetido, la frecuencia de casos

$$C(X1,X2) = \begin{cases} \text{el punto } \_ \ (X_1+X_2) & \text{si } X_1 \neq X_2 \\ \text{el punto } X_1 - 1 & \text{si } X_1 = X_2 \end{cases}$$

donde el conjunto $C(X_1,X_2)$ actualmente contiene θ, converge en probabilidad a 75%. No dice que la probabilidad que $\theta \in C(X_1,X_2)$ es igual a 75% en un caso particular, una interpretación que muchos usuarios dan.

En este ejemplo en particular, es claro que la información contenido en la muestra es muy diferente si $X_1 = X_2$ ó si $X_1 = X_2$. En el primer caso, hemos observado $X_1 = X_2 = \theta - 1$ ó $X_1 = X_2 = \theta + 1$. Entonces podemos concluir que ó $\theta = X_1 - 1$ ó $\theta = X_1 + 1$, aparentemente con igual probabilidad par ambos posibilidades (la función de verosimilitud es igual para los dos posibilidades). Entonces sería mejor reportar que el grado de confianza condicional en la muestra de $C(X_1,X2)$ es 50%. En el otro caso, la muestra da información completa acerca de θ, y parece natural reportar un grado de confianza condicional en a muestra de 100%. Pero este concepto nuevo de un grado de confianza condicional en la muestra no esta definido en la teoría clásica.

La diferencia entre los dos conceptos de grado de confianza usado en el anterior párrafo, es que el grado de confianza clásica es una medida pre-experimental (esta basada en probabilidades que se puede calcular antes de obtener la muestra), mientras el grado de confianza condicional en los datos es una medida post-experimental (esta basada en probabilidades que se puede calcular sólo después de obtener la muestra). La Estadística clásica supone, sin decirlo, que una medida pre-experimental es relevante también después de obtener la muestra. Esto está abierto a discusión. El ejemplo anterior muestra que este supuesto por lo menos es dudable.

Si el ejemplo parece artificial, en que algunas muestras posibles revelan con certeza la identidad del parámetro desconocido, existen dos defensas. Primero, era necesario usar un ejemplo lo más sencillo posible para hacer un punto. Segundo, éste fenómeno ocurre en ejemplos prácticos. Considere una pareja, un hombre con ojos azules, seguramente un homozygoto (bb), y una mujer con ojos negros, que puede ser un heterozygoto ó un homozygoto (bB ó BB). Pero después de observar que uno de los hijos de la pareja tiene ojos azules, sabemos que la mujer es heterozygoto. Entonces, casos donde la información contenida en una muestra depende de la configuración de esta, ocurren también en la práctica.

Kiefer [5] ha tratado de desarrollar una teoría de confianza condicional, pero existe muy pocos artículos que han seguido esa línea, lo que implica que ésta teoría sería muy difícil. También es conocido [ver ejemplos en (2)] que ésta teoría no puede ser completamente general.

3. Verosimilitud

Ha llegado el momento para presentar un método que esta propuesto como alternativa a la teoría clásica. Discutimos la propuesta de realizar inferencias basado solo en la función de verosimilitud, sin ningún otro input. Primero tenemos que observar que la función de verosimilitud es fundamental también en otros acercamientos a la estadística, lo que en cualquier caso hace interesante ver que inferencias se puede hacer solo usando esta función.

Si los dados $X = (X_1,X_2,.....X_n)$ tiene función de densidad

$$f(x \mid \theta), \ \theta \in \Theta,$$

Donde = es el espacio parámetrico, definimos la función de verosimilitud como

$$L(\theta \mid x) = k \ \text{x} \ f(x \mid \theta),$$

Omitiendo constantes irrelevantes. Queremos estudiar el "contenido" de soporte para los intervalos de confianza usuales para u, con grado de confianza 95%. Esto tiene la forma.

$$\overline{xn} - t_{0.025,\,n-1}\, s/\sqrt{n},\ \overline{xn} + t_{0.025,\,n-1}\, s/\sqrt{n})$$

y la diferencia en soporte entre el estimador máximo verosímil de u y los puntos extremos de este intervalo de confianza esta dado por

$$\frac{n}{2}\log\left(\frac{1 + t^2_{.025,\,n-1}}{n-1}\right)$$

Una pequeña tabla de estos valores esta dado en la continuación:

n	Diff en soporte	Razón de verosimilitud
2	5.09	162.4
3	3.49	32.8
4	2.95	19.1
5	2.68	14.6
6	2.53	12.6
10	2.25	9.5
20	2.07	7.92
1001	1.92	6.82

El primer resultado, obvio, de este ejercicio, es que el contenido evidencial de los intérvalos de confianza basado en la distribución t depende de los grados de libertad, y segundo, que para pocos grados de libertad son demasiados conservativos. Esta observación es consistente con la paradoja de Lindley [mire(l)] en la teoría Bayesiana: El contenido evidencial de una prueba clásica he hipótesis depende de los números de observaciones n.

Como un compromiso con la teoría de verosimilitud pura, sería conveniente siempre reportar las diferencias en soporte cuando se da un intervalo de confianza.

Damos un ejemplo más, esta vez la prueba *F* en el modelo lineal general. Suponemos $Y = X\beta + \epsilon$, con los supuestos estándares de normalidad e independencia. Estamos interesados en la hipótesis $H : A\beta = c$. La prueba *F* equivalente con la prueba de razón de verosimilitud, esta dado por

$$F = \frac{(A\beta - c)^T [A(X^T X)^{-1} A^T]^{-1} (A\beta - c)}{q\, S^2}$$

Donde q es el número de restricciones lineales, en el hipótesis H, y S^2 es el estimador usual (insesgado) de la varianza. Bajo la hipótesis nula *F* tiene una distribución F con *q y n-p* grados de libertad. Algunos cálculos sencillos basados en formulas dados en por ejemplo Seber [8], muestra que la diferencia en soporte ente el estimador máximo verosímil irrestringido y bajo la hipótesis H, esta dado por

$$\frac{n}{2}\log\left(1 + \frac{q}{n-p}F\right)$$

lo que en el caso t se reduce a la expresión dado anteriormente.

Usamos esto para el análisis del ejemplo de penicilina de Box, Hunter & Hunter [3]. Este ejemplo es un diseño de bloques al azar para comparar 4 diferentes métodos para la producción de penicilina. Para este ejemplo, $n = 20$, $n - p = 12$. En el cuadro abajo damos los datos relevantes, calculados a partir del tabla de ANOVA.

Hipótesis	q	F	Diff en soporte	Razón de verosimilitud
Tratamientos	3	1.24	2.7	14.8
Bloques	4	3.51	7.75	2315

Pero aquí estamos comparando modelos con mayor diferencia en complejidad, es decir, número de parámetros, entonces debemos requerir mayor diferencia en soporte para declarar una diferencia como evidencia fuerte. Edwards [4] note que se tome una diferencia en soporte de dos unidades como un buen taza de intercambio entre evidencia y complejidad cuando se compara dos modelos con un parámetro en diferencia, entonces en general se puede requerir dos unidades de soporte par cada parámetro en diferencia. En este criterio la insignificancia de tratamientos es evidente, mientras la significancia de bloques parece claro. En la teoría clásico, la prueba *F* del hipótesis nula de ningún efecto de los bloques da un valor - $p \approx 0.04$, que tradicionalmente se juzga como significante. Al respecto, Box, Hunter & Nunter [3] dice: donde k es una constante positivo arbitraría. Esta definición implica que la función de verosimilitud, en isolación, solo se puede usar para hacer inferencias relativas, comparando diferentes posibles valores de θ en los mismos datos. Esto es, los proponentes de verosimilitud puro niega la posibilidad de hacer inferencias absolutas, como propone la Estadística clásica ó Estadística Bayesiana, dice que solo inferencias relativas son posibles. El requerimiento de que evidencia en conjuntos de datos independientes, acerca del mismo parámetro θ, debe ser aditiva, llega a medir la evidencia usando la función de log verosimilitud, llamado la función de soporte en Edwards [4]. El valor de θ con mayor soporte corresponde al estimador de máximo verosimilitud, pero clásicamente se ha evaluado el estimador de máximo verosimilitud, pero clásicamente se ha evaluado el estimador de máximo verosimilitud, usando sus propiedades en muestreo repetido, mientras esto no es permitido (ó tiene sentido) en la teoría de verosimilitud puro. Interpretamos la razón de verosimilitud entre dos diferentes hipótesis (el exponencial de la diferencia en soporte) directamente como una medida de la evidencia relativa entre los dos hipótesis. En Estadística clásica se niega directamente que esto tiene sentido, aunque en libros modernos no tan fuerte como antes.

Edwards en [4] propone, usando una comparación con inferencia clásica en el modelo normal con varianza conocida, usar una diferencia en soporte de 2 unidades como un criterio de "plausibilidad". Es decir, propone que todos los valores de θ tiene diferencia en soporte comparado con el estimador máximo verosímil menor que 2, como plausibles. Si la diferencia en número de parámetros entre los dos hipótesis son mayor a uno, es necesario de alguna manera compensar más por la diferencia en complejidad entre los dos hipótesis. Porque, todo otro igual preferimos la hipótesis más simple que es consistente con los hechos.

Vamos a estudiar el criterio de diferencia en soporte (ó equivalentemente, razón de verosimilitud), usándolo en algunos modelos simples bien conocidos.

Primero el modelo normal con esperanza y varianza desconocida. Sea

$$X_1, X_2, \ldots\ldots, X_n \text{ iid} \quad N(\mu, \sigma^2)$$

La función de prueba $T_n = (x_n - \mu)/ s\sqrt{n}$ tiene la distribución tn - 1 , y da lugar a una

Función de soporte proporcional a

$$l(t) = -\frac{n}{2} \log\left(1 + \frac{t^2}{n-1}\right),$$

A menudo se confunde el contraste F general . Un analista que encuentra un valor significativo a nivel de 4% podría ir directamente a la conclusión de que había mucho más evidencia para la separación entre las medias que la realmente hay. Sin embargo, el examen de la figura 7.5.b le haría ver la realidad de que aunque es probable que haya diferencias de alguna clase entre las medias (por ejemplo, mezcla 1 y 5 probablemente tienen medias diferentes), no están bien calculadas. En concreto, si las diferencias de, por ejemplo, 5 unidades fueran de importancia económica, haría falta un experimento mucho más grande para estimarlas.

(La figura 7.5.b referida es una distribución de referencia para diferencias entre las promedias) . Usando una lasa de intercambio de dos unidades de soporte por parámetro, el análisis de verosimilitud puro da una conclusión parecida a la conclusión de Box, Hunter & Hunter 13) . La evidencia para un efecto importante de bloques es débil. El uso automático de la prueba F puede fácilmente llegar a una conclusión errónea.

Incidentalmente , la penalización de dos unidades de soporte por parámetro indicado por Edwards [4) , es más fuerte que el implicado en Akaikes criterio AIC:

AIC = - 2 log (verosimilitud maximizado) - 2 q,

Que corresponde a una penalización de uno por parámetro. Usando este criterio bloques son significativos en el ejemplo. El criterio AIC ha sido muy popular en ciencias aplicadas, especialmente en tecnología, los últimos décadas. Es probablemente la instancia más usado de inferencia de verosimilitud pura.

4. Conclusiones y comentarios.

Hemos visto que inferencia de verosimilitud puro es un punto de vista interesante que puede dar resultado diferentes de Estadística Clásicas y en muchos casos puede ser por lo menos un complemento importante. En las referencias, especialmente [2] y [4] se pude encontrar muchos más ejemplos.

En la enseñanza de Estadística se debe presentar los diferentes puntos de vista en el conflicto acerca de inferencia Estadística, y especialmente mostrar que medidas pre-experimentales de la Estadística clásica puede ser inválidos, desde un punto de vista post-experimental. Los dos puntos de vista más importantes como alternativas a la Estadística clásica, Bayes y verosimilitud puro, ambos usa un punto de vista post-experimental. Se presenta este papel como una invitación a una discusión aquí, localmente, acerca de estos problemas.

Referencias

(*1) James O. Berger (1980) Statistical Decisión Theory: Foundations, Concepts, and Methods. Springer - Verlag.*

(1) James O. Berger A Robert L Wolpert (1984) The Likelihood Principie, Institute of Marthemathical Statistics, Lecture Notes - Monograph Series.

(2) G.E.P. Box, W. G. Hunter A J.S. Hunter Statistics for Experimenters, and Introduction to Design, Data Analysis and Model Building. John Wiley A Sons.

(3) A. W.F. Edwards (1992) Likelihood. Expanded Edition. The John Hopkins University Press.

(4) J.Kiefer (1977) Conditional confidence statements and confidence estimators (with discussion) J. Amer. Statist. Assoc.. 72, 789-827.

(5) Hacking (1979) Logic of Statistical Inference. Cambridge University Press.

(6) Kjetil Halvorsen (1997) ¿Es probabilidad un concepto suficiente para expresar incertidumbre? Apuntes de enseñanza.

(7) G.A.F. Seber (1977) Linear Regressión Analysis. Jhon Wiley A Sons.