



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS MATEMÁTICAS Y DE LA
ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA

CONSTRUCCIÓN DE INTERVALOS: USO DE PIVOTALES, INFERENCIA FIDUCIAL Y USO DE LA DISTRIBUCIÓN *A-POSTERIORI*

T E S I S

QUE PARA OBTENER EL GRADO DE:

**MAESTRA EN ESTADÍSTICA E
INVESTIGACIÓN DE OPERACIONES**

P R E S E N T A :

MARÍA DEL PILAR ALONSO REYES

T U T O R :

DR. FEDERICO JORGE O´REILLY TOGNO



2009

Agradecimientos

A mi jurado:

A Federico O'Reilly. Porque siempre tuvo confianza en mi y me dio apoyo incondicional, tanto como maestro como director de este trabajo.

Silvia Ruiz-Velasco Acosta, Luis Antonio Rincón Solís, Juan González Hernández y José Antonio Flores Díaz. Porque me aportaron mucho con sus correcciones y observaciones e hicieron que este trabajo se presente con una gran calidad.

A mis amigos

Mirna Ocampo. Gran amiga que hizo posible que este trabajo se realizara, apoyándome siempre con la mecanografía.

Gerardo Chávez. Por su apoyo teórico y consejos computacionales.

Luis Guillermo Cota. Por su apoyo computacional.

Dedicatorias

A mi madre María del Pilar Reyes Ferra †, porque siempre fuiste y serás un impulso muy importante en mi vida.

A mi padre Enrique Alonso Colmenares, porque sigues dándome apoyo en todo lo que emprendo.

A Alfredo Pérez Alonso porque eres la parte esencial de mi vida.

Índice

Introducción	I
Capítulo 1	1
Estimación por intervalos: uso de pivotaes	1
1.1. Información y suficiencia	1
1.1.1. Función de verosimilitud	1
1.1.2. Suficiencia	2
1.1.3. Información de Fisher y estadística suficiente	5
1.1.4. Principios de inferencia	6
1.2. Estimación por intervalos	8
1.2.1. Regla empírica	8
1.3. Intervalos de confianza	9
1.4. Procedimientos de construcción de intervalos usando la cantidad pivotal	12
1.4.1. Cantidad pivotal.....	12
1.4.2. Procedimiento de la cantidad pivotal	14
1.4.3. Uso explícito de la estadística suficiente minimal.....	20
1.4.4. Importancia de los valores q_1 y q_2	25
1.4.5. Aplicación del procedimiento pivotal con uso de la cantidad pivotal asintótica.....	25
Capítulo 2	34
Argumento fiducial y enfoque Bayesiano.....	34
2.1. Argumento fiducial.....	34

2.1.1. Argumento fiducial cuando es explícita la función de distribución de la estadística suficiente minimal.....	39
2.1.2. Argumento fiducial implícito	46
2.2. Enfoque Bayesiano	53
Capítulo 3	66
Distribuciones fiducial y posterior. Semejanzas y diferencias	66
3.1. Características básicas de la distribución fiducial.....	66
3.2. Distribución fiducial	68
3.3. Revisión del cumplimiento de la contrastabilidad monótona	69
3.4. Ejemplos donde el parámetro no es de localización o hay masa en alguno de los extremos del intervalo del parámetro.	71
Conclusiones	88
Bibliografía.....	90

Introducción

Uno de los temas importantes en la inferencia estadística, es el problema de la estimación por intervalos. El procedimiento más visto y desarrollado en los libros de estadística es el de la construcción de intervalos de confianza basados en la cantidad pivotal; ignorando que existen otros procedimientos y enfoques que permiten de igual manera hacer inferencia vía intervalo de uno o varios parámetros.

Uno de los objetivos de este trabajo de tesis es el de mostrar un panorama de los enfoques clásico, fiducial y Bayesiano para el cálculo de intervalos de confianza, intervalos fiduciales y los de más alta probabilidad final, respectivamente. Siguiendo estas construcciones, el siguiente objetivo radica en que el lector vea cómo las funciones de densidad fiducial y las funciones finales coinciden para algunos ejemplos, bajo ciertas condiciones; y como último objetivo es mostrar que cuando las condiciones no se cumplen, los enfoques fiducial y Bayesiano algebraicamente arrojarán resultados diferentes, pero numéricamente muy parecidos.

En cada capítulo se muestra la teoría relacionada al tema propuesto y se ilustra con ejemplos para hacer una mejor discusión del mismo, así en el capítulo I se proporciona la herramienta básica para la estimación por intervalos con el punto de vista de la estadística clásica, revisando la utilización de cantidades pivotaes tanto exactas como asintóticas terminando con resultados asintóticos que Sprott (1975) desarrolló para la estimación por intervalos.

En el capítulo II se revisa el tema de estimación por intervalo desde los puntos de vista

fiducial y Bayesiano. En una primera parte se inspeccionan las nociones básicas del enfoque fiducial basado en la utilización de cantidades fiduciales y también se dan los conceptos necesarios de este procedimiento para cuando no se hace uso explícito de cantidades pivotaes. En la segunda parte se exponen de manera abreviada el enfoque Bayesiano para la construcción de intervalos de probabilidad y se hace una comparación de los tres, a través de ejemplos.

En el capítulo III se hace un resumen de las características básicas de las distribuciones fiduciales y se dan los razonamientos de contrastabilidad para luego, a través de ejemplos, revisar que lo que indica Lindley (1961) de que existe coincidencia en las distribuciones fiducial y alguna posterior bajo ciertas condiciones en términos de transformaciones en los parámetros; pero también se hace hincapié a través de ejemplos que cuando se viola una de las consideraciones de Lindley, efectivamente se tendrán diferentes expresiones para las distribuciones fiducial y posterior, pero su parecido numérico extraordinario.

Por último se presentan las conclusiones y la bibliografía que sustenta este trabajo.

Capítulo 1

Estimación por intervalos: uso de pivotaes

Introducción

En este capítulo se hará una revisión somera de la función de verosimilitud, de las estadísticas suficientes y su relación con la información de Fisher, con el fin de mostrar su aplicación en la estimación por intervalos a través del procedimiento de la cantidad pivotal; con diversos ejemplos se muestra su uso y sus características.

1.1. Información y suficiencia

1.1.1. Función de verosimilitud

Para construir un intervalo de confianza de un parámetro θ de una función de densidad, es necesario disponer de mecanismos que permitan relacionar la información muestral con dicho parámetro y una vía es la función de verosimilitud.

Si se considera una muestra aleatoria, entonces todos y cada uno de sus elementos u observaciones se suponen son variables aleatorias independientes (v.a.i.) y con la misma función de densidad de probabilidad que la población. Como en esta última se desconoce el verdadero valor del parámetro θ , cada uno de los elementos de la muestra presentará una función de densidad de probabilidad que también depende de θ , por lo que la función de densidad de probabilidad conjunta de la muestra es:

$$f(x_1, \dots, x_n; \theta) = f(x_1; \theta) \cdots f(x_n; \theta)$$

Esta función conjunta se denota por $f(\underline{x}; \theta)$.

Se tienen dos consideraciones importantes respecto a esta función conjunta:

- a) La muestra observada, \underline{x} , se considera variable y θ toma un valor fijo, aunque desconocido, del espacio parametral Θ , entonces $f(\underline{x}; \theta)$ es la función de densidad de probabilidad conjunta de la muestra, evaluada en \underline{x} .

- b) Por otro lado al considerar que la muestra \underline{x} es fija y el parámetro θ puede tomar cualquier valor en Θ indica que $f(\underline{x}; \theta)$, ahora denotada por $L(\theta; \underline{x})$, sea sólo función del parámetro y lleva al concepto de verosimilitud, como una medida de la credibilidad de los distintos valores del parámetro, esto quiere decir que la verosimilitud en realidad informa sobre la "plausibilidad" de ciertos valores de θ relativos a otros valores de θ y por ello no tiene una escala definida. Por lo anterior suele considerarse la función de verosimilitud como $L(\theta; \underline{x}) \propto f(\theta; \underline{x})$.

1.1.2. Suficiencia

Para hacer inferencias sobre el vector θ se parte de la información que pueda suministrar la muestra aleatoria, $\underline{X} = (X_1, \dots, X_n)$, resumiéndose en una estadística $S(\underline{X})$.

Al resumir la información en $S(\underline{X})$, se puede preguntar si no se pierde información acerca de θ . La respuesta la da Fisher (1922) mostrando que una *estadística es suficiente* para efectuar inferencias sobre θ cuando logra resumir el contenido de

información relevante suministrada por la muestra. Esto es, habiendo observado $S(\underline{X})$, ninguna otra estadística, función de \underline{X} , puede proporcionar *información adicional* acerca de θ .

Lo anterior significa que si una estadística $S(\underline{X})$ es *suficiente*, el conocimiento minucioso de los elementos muestrales no añade nada más sobre θ que no hubiera proporcionado $S(\underline{X})$. Esto es lo mismo que decir que la distribución de \underline{X} condicionada a que la estadística $S(\underline{X}) = s$ no dependa de θ ; si sucediera lo contrario, indicaría que al analizar detalladamente a \underline{X} , aún conociendo $S(\underline{X}) = s$, se tendría información adicional de θ .

Fisher proporciona un criterio sencillo para comprobar la suficiencia de una estadística, que es el siguiente:

Teorema 1.1. Factorización. (Estadística suficiente)

Una condición necesaria y suficiente para que la estadística $S(\underline{X})$ sea suficiente es que la función de verosimilitud de la muestra, $L(\theta; \underline{x})$ se descomponga como el producto de dos funciones; una $g(s(\underline{x}); \theta)$ que depende de \underline{x} a través de la estadística $s(\underline{x})$, y del parámetro; y otra función, $H(\underline{x})$, que es independiente de θ , es decir:

$$L(\theta; \underline{x}) = g(s(\underline{x}); \theta) \cdot H(\underline{x}) \quad (1)$$

Un resultado importante sobre suficiencia es el siguiente:

Teorema 1.2.

Un estimador máximo verosímil o un conjunto de estimadores máximo verosímiles dependen de la muestra a través de un conjunto de estadísticos suficientes.

El resultado se deriva de la aplicación del teorema de factorización, ya que si $S = s(\underline{X})$ es suficiente, entonces la función de verosimilitud puede escribirse como:

$$\begin{aligned} L(\theta; \underline{x}) &= \prod_{i=1}^n f(x_i; \theta) \\ &= g(s(\underline{x}); \theta) \cdot H(\underline{x}) \end{aligned}$$

Como una función de θ , $L(\theta; \underline{x})$, tendrá su máximo en el mismo lugar que $g(s(\underline{x}); \theta)$; pero este lugar donde g se maximiza depende de \underline{x} sólo a través de s .

Como ya se refirió anteriormente, el concepto de suficiencia permite condensar la información que proporciona la muestra, sin perder la que se tiene sobre el parámetro, pero puede haber más de un conjunto de estadísticas suficientes para el parámetro en cuestión.

Si se toma una muestra de una población normal con media y varianza desconocida, se puede ver que, al menos hay tres estadísticas suficientes para los parámetros:

- i) La muestra X_1, \dots, X_n en sí misma es suficiente, obviamente;
- ii) Las estadísticas de orden $Y_1 = \text{mín}\{X_1, \dots, X_n\}, \dots, Y_n = \text{máx}\{X_1, \dots, X_n\}$ también constituyen un vector suficiente, y
- iii) Finalmente, el vector conocido con dos componentes $\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2$ resulta también suficiente¹.

¹ Cabe señalar que se pueden tomar funciones de ellas y generar otras estadísticas suficientes como: \bar{X} y

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}.$$

Entonces, se buscaría de entre las posibles estadísticas suficientes, la que más condense la información que se tiene. Ello conlleva al concepto de lo que se conoce como *estadística suficiente minimal*, que para este ejemplo resulta ser $\sum_{i=1}^n X_i$ y $\sum_{i=1}^n X_i^2$.

Esta condensación de la información del parámetro se da en el concepto de estadística suficiente minimal, por lo que ésta se define como:

Definición 1.1. Estadística suficiente minimal

Una estadística suficiente se define como suficiente minimal si y sólo si es una función de cualquier otra estadística suficiente.

1.1.3. Información de Fisher y estadística suficiente

Toda la información disponible sobre un parámetro está contenida en la muestra. Fisher indica que una buena medida de qué tanta información existe, está dada por el cambio que muestre la verosimilitud ante variaciones del parámetro. El tamaño del cambio de $L(\theta; \underline{x})$ ante variaciones del parámetro θ indicaría cuánta información sobre dicho parámetro contiene la muestra.

Así, Fisher define una cantidad aleatoria, el *score* o puntaje como:

$$\frac{\partial \log L(\theta; \underline{X})}{\partial \theta}, \tag{2}$$

el cual recoge los movimientos de la función de verosimilitud ante cambios infinitesimales de θ .

Para medir la información que la muestra proporciona sobre θ , Fisher consideró la matriz de varianza-covarianza del score y la llamó información ²:

$$I(\theta) = -E\left(\frac{\partial^2 \log L(\theta; \underline{X})}{\partial \theta_i \partial \theta_j}\right) \quad (3)$$

La información suministrada por una estadística $S(\underline{X})$ es siempre igual o menor a la dada por la muestra si y sólo si $S(\underline{X})$ es suficiente conteniendo la misma información de la muestra.

1.1.4. Principios de inferencia

Principio de suficiencia

Si la estadística $S(\underline{X})$ es suficiente para el parámetro θ , y $S(\underline{x}_1) = S(\underline{x}_2)$, entonces la misma inferencia puede ser considerada si se observa \underline{x}_1 o \underline{x}_2 .

Principio de verosimilitud

Sean \underline{x}_1 y \underline{x}_2 dos posibles muestras de $f(x; \theta)$, si las verosimilitudes son "las mismas", esto es, si

$$L(\theta; \underline{x}_1) \propto L(\theta; \underline{x}_2) \quad (4)$$

entonces la misma inferencia debe ser obtenida con \underline{x}_1 o con \underline{x}_2 .

² Para obtener la varianza del score deben cumplirse las condiciones de regularidad de Fisher-Wolfowitz:

1. La población de donde procede la muestra tiene un campo de variación que no depende de θ .
2. La función $L(\theta; \underline{x})$ tiene al menos las dos primeras derivadas respecto a θ .
3. Las operaciones de derivación e integración (o suma en el caso discreto) son intercambiables.

Principio fuerte de verosimilitud

Considere dos muestreos en la que aparece el mismo parámetro θ . En el primero se observa una muestra \underline{x} y se evalúa la verosimilitud $L_1(\theta; \underline{x})$. En el segundo se observa \underline{y} y la verosimilitud es $L_2(\theta; \underline{y})$. Si $L_1(\theta; \underline{x}) \propto L_2(\theta; \underline{y})$, entonces las inferencias sobre θ utilizando \underline{x} o \underline{y} de su correspondiente esquema, deben ser idénticas.

Existe la noción de ***equivalencia en verosimilitud*** en la que para $\underline{x}_1, \underline{x}_2$ si $L(\theta; \underline{x}_1) \propto L(\theta; \underline{x}_2)$, se dice que \underline{x}_1 es equivalente en verosimilitud a \underline{x}_2 . Esta definición presupone que las posibles x 's pertenecen al soporte³ de la distribución. Se dice que \underline{x} está en el soporte si existe θ para el cual $L(\theta; \underline{x}) > 0$. Obsérvese que si se eliminan las x 's que no están en el soporte, no significa nada, pues una \underline{x} que no esté en el soporte equivale a que $f(\underline{x}; \theta) = 0, \forall \theta$.

La suficiencia minimal de una estadística suficiente S puede constatarse, si es que, cada vez que \underline{x}_1 es equivalente en verosimilitud a \underline{x}_2 (ambas en el soporte) se sigue necesariamente que $S(\underline{x}_1) = S(\underline{x}_2)$.

La equivalencia en verosimilitud produce una partición en el espacio muestral (intersectado con el soporte), dada por las clases de equivalencia que corresponden a la partición que induce una estadística suficiente minimal. Resulta, entonces, inmediato que cualquier función uno a uno de la estadística suficiente minimal es suficiente minimal.

³ El soporte es entendido como el dominio donde toman sus valores las x .

1.2. Estimación por intervalos

En inferencia estadística suele hacerse la llamada estimación puntual, que es construir un estimador (función de la muestra) que dé una idea del valor paramétrico de interés. Desde luego que como cantidad aleatoria, el estimador fluctuará y se buscará que “oscile poco” alrededor del verdadero valor.

Existe como noción en la inferencia estadística, el llamado “error de estimación” que no es más que la desviación estándar del estimador, y con ello se estila utilizar la llamada “regla empírica”. Ésta proporciona un procedimiento para apreciar un intervalo y presupone muchas cosas sobre la población y la muestra, que no siempre pueden cumplirse, como lo es el tener una muestra grande que garantice normalidad del estimador, al menos aproximadamente. Existen desde luego otros procedimientos para realizar una estimación por intervalos aparte de la regla empírica que serán revisados en este capítulo.

1.2.1. Regla empírica

De la estimación puntual se puede inferir por medio de la regla empírica y construir un intervalo que permita asegurar:

- 1) Es “probable” que $\theta \in \left(\hat{\theta} - \sqrt{\widehat{\text{var}}(\hat{\theta})}, \hat{\theta} + \sqrt{\widehat{\text{var}}(\hat{\theta})} \right)$
- 2) Es “muy probable” que $\theta \in \left(\hat{\theta} - 2\sqrt{\widehat{\text{var}}(\hat{\theta})}, \hat{\theta} + 2\sqrt{\widehat{\text{var}}(\hat{\theta})} \right)$
- 3) “Casi seguramente” se tiene que $\theta \in \left(\hat{\theta} - 3\sqrt{\widehat{\text{var}}(\hat{\theta})}, \hat{\theta} + 3\sqrt{\widehat{\text{var}}(\hat{\theta})} \right)$

donde $\hat{\theta}$ es el estimador puntual de θ y $\widehat{\text{var}}(\hat{\theta})$ representa la varianza estimada de $\hat{\theta}$.

Es importante señalar que en los incisos anteriores los coeficientes que acompañan a la $\sqrt{\text{var}(\hat{\theta})}$ (1, 2, y 3 respectivamente) corresponden al 64, 95 y 99% aproximadamente de una distribución normal.

El problema de utilizar la regla empírica es que se necesita que $\hat{\theta}$ tenga una distribución aproximadamente normal. Cuando hay colas pesadas o asimétricas en la distribución se obtendrían con este procedimiento intervalos erróneos.

Una aparente salida a este problema es la consideración muestras "grandes" que permitan proponer estimadores que posean las cualidades de consistencia y normalidad, lo cual garantizaría una situación apropiada para la aplicación de la regla empírica.

Desafortunadamente no es sencillo en cada caso concreto poder aseverar que la muestra es "grande". En algunos casos $n = 100,000$ aún no es "grande". Por ejemplo si se toma una muestra con $n=7$ de una función Bernoulli⁴ con parámetro (supuestamente desconocido) con valor cercano a 0.5, es asombroso lo bien que aproxima a la verdadera distribución de \bar{X} (media muestral), una distribución normal. Sin embargo si el valor del parámetro fuese, por ejemplo 0.001, la verdadera distribución de \bar{X} es muy distinta de la considerada anteriormente.

1.3. Intervalos de confianza

Con las consideraciones del apartado anterior sobre la estimación puntual, se busca

⁴ $f(x; p) = p^x (1-p)^{1-x}$; $x = 0,1$

ahora, proponer una forma alternativa de conocer el parámetro θ de una función $f(x;\theta)$, utilizando únicamente la muestra, x_1, \dots, x_n y la distribución asociada a ella. Dicha propuesta contempla la estimación por intervalos.

Los *intervalos de confianza* pueden definirse como subconjuntos aleatorios del espacio parametral Θ , en donde la probabilidad de que contengan al verdadero valor del parámetro θ es $\gamma = 1 - \alpha$.

A continuación se dará una definición formal de un intervalo de confianza y después de región de confianza.

Definición 1.2. Intervalo de Confianza

Sea el vector $\underline{X} = (X_1, \dots, X_n)$ de variables aleatorias con función de distribución conjunta dependiente del parámetro θ y sean $L(\underline{X})$ y $U(\underline{X})$ dos estadísticas, con $L(\underline{x}) < U(\underline{x})$. El intervalo, habiendo observado $\underline{X} = \underline{x} = (x_1, \dots, x_n)$, $[L(\underline{x}), U(\underline{x})]$, es llamado intervalo al $(1 - \alpha) \times 100\%$ de confianza para θ si:

$$P_{\theta} [L(\underline{X}) \leq \theta \leq U(\underline{X})] = 1 - \alpha . \tag{5}$$

Definición 1.3. Región de confianza

Sea $\underline{X} = (X_1, \dots, X_n)$ un vector de variables aleatorias con función de distribución conjunta dependiente de un vector de parámetros $\theta \in \Theta$ y sea $R(\underline{X})$ un subconjunto de

Θ , dependiente de \underline{x} y por tanto aleatorio. Entonces $R(\underline{x})$ es llamada región al $(1 - \alpha) \times 100\%$ de confianza para θ si:

$$P_{\theta}[\theta \in R(\underline{X})] = 1 - \alpha. \quad (6)$$

Se pueden obtener diversos intervalos o regiones, según sea el caso, por lo que debe haber criterios para comparar unos u otros. Un criterio de optimalidad es lograr la región que sea estocásticamente más pequeña, es decir, dado algún valor pequeño α y dos regiones de confianza distintas, por ejemplo $R_1(\underline{x})$ y $R_2(\underline{x})$ para el parámetro θ , se dice que $R_1(\underline{x})$ es preferible a $R_2(\underline{x})$ si se satisface lo siguiente:

$$P(\theta' \in R_1(\underline{X}); \theta) \leq P(\theta' \in R_2(\underline{X}); \theta) \quad \forall \theta' \neq \theta \in \Theta$$

Lo que indica que $R_1(\underline{X})$ es estocásticamente más pequeña que $R_2(\underline{X})$, si la confianza de que $R_1(\underline{x})$ contenga un valor parametral equivocado θ' , es menor o igual que la confianza de que $R_2(\underline{x})$ lo contenga.

Hay otros requisitos para la construcción de regiones de confianza que ya no se discuten aquí, como lo es el que exista "anidamiento", ver por ejemplo a Aznar (1984).

1.4. Procedimientos de construcción de intervalos usando la cantidad pivotal

1.4.1. Cantidad pivotal

Para la construcción de regiones o intervalos de confianza pueden usarse cantidades aleatorias que dependen tanto de la muestra \underline{X} como del parámetro θ y cuya distribución es única, a esto se le conoce como *cantidad pivotal exacta*.

Definición 1.4. Cantidad pivotal exacta

Sea X_1, \dots, X_n una muestra aleatoria de una densidad $f(x; \theta)$. Sea $Q = Q(X_1, \dots, X_n; \theta)$ una función de X_1, \dots, X_n y de θ . Si Q tiene una función de distribución $G(q)$ ($G(q) = P[Q \leq q]$) que no depende de θ , entonces Q es llamada *cantidad pivotal exacta*.

Ejemplo 1.1.

Sean X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas (iid) $Exp(\lambda)$ ($f(x; \lambda) = \lambda e^{-\lambda x}$; $\lambda > 0$), una cantidad pivotal exacta es:

$$Q = \lambda \sum_{i=1}^n X_i$$

que depende solamente de X_1, \dots, X_n y de λ , su función de densidad es gama con parámetros de forma n y de escala 1, es decir, $g(q) = \frac{1}{\Gamma(n)} q^{n-1} e^{-q}$; $q > 0$.

Cuando se apela a la teoría de "muestras grandes", se define una *cantidad pivotal asintótica* como aquella función $Q_n = Q_n(\underline{X}; \theta)$ de la muestra y el parámetro, cuya distribución $G_n(q)$, en el límite es $G(q)$, independiente de θ .

Ejemplo 1.2.

Sean X_1, \dots, X_n variables aleatorias iid con función de densidad de probabilidad:

$$f(x; \theta) = \frac{3x^2}{\theta^3}; \quad 0 \leq x \leq \theta$$

El estimador máximo verosímil de θ es $Y_n = \text{máx}\{X_1, \dots, X_n\}$ y si n es grande, se tiene que la cantidad pivotal asintótica es:

$$Q_n = \frac{(3n+1)Y_n - 3n\theta}{\theta \sqrt{\frac{3n}{3n+2}}} \sim N(0,1)$$

Se han mostrado dos ejemplos fáciles que permiten obtener expresiones relativamente sencillas para la cantidad pivotal; en la práctica esto no siempre se cumple, por lo que a veces no se sabe qué cantidad pivotal utilizar.

Mood, Graybill y Boes (1974) señalan el procedimiento usando la cantidad pivotal y comentan las características que tiene y aclaran que si la estadística suficiente para θ tiene una distribución continua, entonces siempre es posible construir la cantidad pivotal

⁵.

⁵ Si $U(X_1, \dots, X_n)$ es suficiente para θ y $P[U \leq u; \theta] = H(u; \theta)$ es continua, entonces $H(u; \theta) \sim U(0,1)$.

1.4.2. Procedimiento de la cantidad pivotal

Se supone que se tiene X_1, \dots, X_n variables aleatorias iid con función de densidad de probabilidad $f(x; \theta)$ parametrizada por θ , el procedimiento descrito a continuación se puede usar para θ o para $\tau(\theta)$ como función real de θ .

Sea $Q = Q(X_1, \dots, X_n; \theta)$ una cantidad pivotal exacta y para algún valor fijo $\gamma \in (0, 1)$, donde $\gamma = 1 - \alpha$, existirán q_1 y q_2 dependientes de γ tales que:

$$P(q_1 < Q < q_2) = \gamma. \quad (7)$$

Si la distribución de Q es continua, para cualquier γ uno puede obtener q_1 y q_2 .

Si la distribución es discreta hay limitaciones, ya que no se tendría en (7) la igualdad.

Si para cada posible valor muestral x_1, \dots, x_n se tiene $q_1 < Q(x_1, \dots, x_n; \theta) < q_2$
 $\Leftrightarrow t_1(x_1, \dots, x_n) < \tau(\theta) < t_2(x_1, \dots, x_n)$ para funciones t_1 y t_2 (no dependientes de θ),
entonces (t_1, t_2) es el intervalo de confianza al $\gamma \times 100\%$, o bien comúnmente conocido como $(1 - \alpha) \times 100\%$ para $\tau(\theta)$.

Acotaciones con respecto al procedimiento.

1. Para algún valor fijo de γ , hay muchos pares posibles de números q_1 y q_2 que pueden ser seleccionados para que se dé $P(q_1 < Q < q_2) = \gamma$, por lo que

diferentes q_1 y q_2 producen distintos t_i' , en el caso en que el interés sea en $\tau(\theta)$, así que lo deseable es que t_1 y t_2 no estén tan alejados, significando que la longitud del intervalo sea pequeña, es decir:

$$t_2(x_1, \dots, x_n) - t_1(x_1, \dots, x_n) = LI(x_1, \dots, x_n)$$

y entonces, al escoger bien q_1 y q_2 , se tendrá que la longitud del intervalo, $LI(x_1, \dots, x_n)$, sea mínima. Si no es posible que la elección de q_1 y q_2 , que dependen de x_1, \dots, x_n conduzca siempre a que $LI(x_1, \dots, x_n)$ sea mínima, entonces se buscaría que, en promedio $LI(x_1, \dots, x_n)$ sea pequeña.

2. La cantidad pivotal en la desigualdad $q_1 < Q(x_1, \dots, x_n; \theta) < q_2$ puede ser reescrita o pivoteada como: $t_1(x_1, \dots, x_n) < \tau(\theta) < t_2(x_1, \dots, x_n)$.

Ejemplo 1.3.

Tomando los datos del ejemplo 1.1. y la cantidad pivotal calculada, se tiene: X_1, \dots, X_n v.a. iid $Exp(\lambda)$ y $Q = \lambda \sum_{i=1}^n X_i$, el intervalo de confianza se calcularía como:

$$P\left(q_1 \leq \lambda \sum_{i=1}^n X_i \leq q_2\right) = 1 - \alpha, \text{ quedando } P\left(\frac{q_1}{\sum_{i=1}^n X_i} \leq \lambda \leq \frac{q_2}{\sum_{i=1}^n X_i}\right) = 1 - \alpha,$$

donde q_1 y q_2 cuantiles de una función de distribución gama(1,n), denotada $G(\cdot; 1, n)$ pudiéndose escoger por ejemplo $a = G(q_1; 1, n)$ y $b = 1 - G(q_2; 1, n)$ con $a = p\alpha$ y $b = (1-p)\alpha$, se obtiene entonces $a + b = \alpha$, con $p \in (0, 1)$. Pero estos cuantiles no

implican que se tendría el intervalo más corto; si se quiere precisamente el intervalo de longitud más pequeña, entonces se requeriría que la longitud del intervalo se minimizara bajo ciertas condiciones; es decir: $LI(x_1, \dots, x_n) = \frac{1}{\sum_{i=1}^n x_i} (q_2 - q_1)$ sea mínima sujeta a

que $\int_{q_1}^{q_2} \frac{1}{\Gamma(n)} q^{n-1} e^{-q} dq = 1 - \alpha$, es decir, se deben encontrar valores q_1 y q_2 tales que

$q_1^{n-1} e^{-q_1} = q_2^{n-1} e^{-q_2}$ sujeto a que $\int_{q_1}^{q_2} \frac{1}{\Gamma(n)} q^{n-1} e^{-q} dq = 1 - \alpha$, así el intervalo de probabilidad

resultante para la pivotal Q contendrá los valores de q_i con densidad máxima.

Ejemplo 1.4.

Encontrar el intervalo de confianza para la varianza de una $N(\mu, \sigma^2)$

$$\left(f(x; \mu, \sigma^2) = \frac{e^{-\frac{1}{2\sigma^2}(x-\mu)^2}}{\sqrt{2\pi\sigma^2}} \right), \text{ con } \mu \text{ y } \sigma^2 \text{ desconocidas.}$$

Sean X_1, \dots, X_n v. a. iid $N(\mu, \sigma^2)$, entonces $\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi_{(n-1)}^2$ y un intervalo para σ^2 basado en una cantidad pivotal quedaría:

$$P\left(\frac{\sum (x_i - \bar{x})^2}{q_2} \leq \sigma^2 \leq \frac{\sum (x_i - \bar{x})^2}{q_1} \right) = 1 - \alpha$$

El objetivo sería minimizar el intervalo, así que la longitud quedaría expresada por:

$$LI(x_1, \dots, x_n) = \frac{\sum (x_i - \bar{x})^2}{q_1} - \frac{\sum (x_i - \bar{x})^2}{q_2} = \sum (x_i - \bar{x})^2 \left(\frac{1}{q_1} - \frac{1}{q_2} \right)$$

Se trataría de minimizar $h(q_1, q_2) = \sum (x_i - \bar{x})^2 \left(\frac{1}{q_1} - \frac{1}{q_2} \right)$, lo que conduce a minimizar el factor $\frac{1}{q_1} - \frac{1}{q_2}$. Este problema fue visto por Take y Kett (1959), y para que sea mínima esa longitud del intervalo, q_1 y q_2 deben ser cuantiles de una $\chi^2_{(n+3)}$ generando un intervalo para la varianza σ^2 .

Ejemplo 1.5.

Se quiere construir un intervalo para ρ el coeficiente de correlación de la distribución normal divariada, con densidad:

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{1/2}} e^{-\frac{1}{2(1-\rho^2)} \left\{ \left(\frac{x-\mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x-\mu_1}{\sigma_1} \right) \left(\frac{y-\mu_2}{\sigma_2} \right) + \left(\frac{y-\mu_2}{\sigma_2} \right)^2 \right\}}$$

$$-\infty < x, y < \infty; -\infty < \mu_1, \mu_2 < \infty; \sigma_1\sigma_2 > 0; |\rho| < 1$$

σ_i, μ_i, ρ desconocidos.

Reparametrizando, tomando primero a $\rho\sigma_1\sigma_2 = \sigma_{12} \Rightarrow \rho = \frac{\sigma_{12}}{\sigma_1\sigma_2}$, $\sigma_1 = \frac{\sigma_{12}}{\sigma_2\rho}$; $\sigma_2 = \frac{\sigma_{12}}{\sigma_1\rho}$ y

considerando a su estimador, se tiene: $\hat{\sigma}_{12} = \hat{\sigma}_1 \hat{\sigma}_2 \hat{\rho} = \frac{\sum_{i=1}^n \Sigma(x_i - \bar{x})(y_i - \bar{y})}{n}$.

Sustituyendo en la función de densidad conjunta

$$f(x, y) = \frac{1}{2\pi\sigma_{12} \left(\frac{1-\rho^2}{\rho^2} \right)^{1/2}} e^{-\frac{1}{2(1-\rho^2)} \left\{ \rho^2 \sigma_{12}^2 \left(\frac{x-\mu_1}{\sigma_{12}} \right)^2 - 2\rho^3 \sigma_1 \sigma_2 \left(\frac{x-\mu_1}{\sigma_{12}} \right) \left(\frac{y-\mu_2}{\sigma_{12}} \right) + \sigma_{12}^2 \rho^2 \left(\frac{y-\mu_2}{\sigma_{12}} \right)^2 \right\}}$$

Existe en la literatura la llamada pivotal de Fieller (1954), que está dada por $\bar{Y} - \rho\bar{X}$,

$$\bar{Z} = \bar{Y} - \rho\bar{X} \sim N\left(0, \frac{1}{n}(\sigma_2^2 - 2\rho\sigma_{12} + \rho^2\sigma_1^2)\right) = N(0, \sigma_Z^2)$$

$$\sigma_Z^2 = \frac{1}{n}(\sigma_2^2 - 2\rho\sigma_{12} + \rho^2\sigma_1^2)$$

$$\widehat{Var}(\bar{Z}) = \frac{1}{n}(S_2^2 - 2\rho S_{12} + \rho^2 S_1^2)$$

$$\frac{\widehat{Var}(\bar{Z})}{Var(\bar{Z})} \sim \chi_{(n-1)}^2$$

Se construye la estadística; y por lo tanto la cantidad pivotal,

$$\begin{aligned} \left(\frac{\bar{Z}}{\sqrt{Var(\bar{Z})}}\right)^2 &\sim \chi_{(1)}^2 & \Rightarrow \frac{\frac{\bar{Z}^2}{Var(\bar{Z})}}{\frac{1}{Var(\bar{Z})}} &\sim F_{(1, n-1)} \\ & & & \frac{Var(\bar{Z})}{n-1} \\ & \Rightarrow \frac{\frac{\bar{Z}^2}{Var(\bar{Z})}}{n-1} &\sim F_{(1, n-1)} \end{aligned}$$

Tomando el intervalo al $(1 - \alpha) \times 100\%$ de confianza:

$$P\left(\frac{\frac{\bar{Z}^2}{S_2^2 - 2\rho S_{12} + \rho^2 S_1^2}}{n-1} \leq q\right) = 1 - \alpha$$

donde $q \sim F_{(1, n-1)}^{1-\alpha}$ o bien $\sqrt{q} \sim t_{(n-1)}^{1-\alpha}$.

Trabajando la desigualdad del intervalo se llega a:

$$\rho^2 \left(\bar{x}^2 - \frac{qS_1^2}{n-1} \right) - 2\rho \left(\overline{xy} - \frac{qS_{12}}{n-1} \right) + \bar{y}^2 - \frac{qS_2^2}{n-1} \leq 0.$$

Renombrando los coeficientes de la expresión anterior:

$$A = \bar{x}^2 - \frac{qS_1^2}{n-1}, \quad B = \left(\overline{xy} - \frac{qS_{12}}{n-1} \right), \quad C = \bar{y}^2 - \frac{qS_2^2}{n-1}$$

la desigualdad queda como:

$$A\rho^2 - 2B\rho + C \leq 0$$

o lo que es lo mismo

$$\frac{-\sqrt{B^2 - AC} + B}{A} \leq \rho \leq \frac{\sqrt{B^2 - AC} + B}{A}$$

Resolviendo los casos

1. $B^2 - AC = 0 \Rightarrow \rho = \frac{B}{A}$.
2. Si $B^2 - AC > 0 \Rightarrow B^2 > AC$, hay dos raíces reales diferentes.
3. Si $B^2 - AC < 0$ no hay raíces reales y no existe el intervalo.

Si se tiene además que $\rho \doteq \frac{B}{A}$, entonces:

$$\frac{B}{A} = \frac{\overline{xy} - \frac{aS_{12}}{n-1}}{\bar{x}^2 - \frac{aS_1^2}{n-1}} = \frac{(n-1)\overline{xy} - aS_{12}}{(n-1)\bar{x}^2 - aS_1^2}$$

si se considera que $\frac{\bar{y}}{\bar{x}} \doteq \frac{S_{12}}{S_1^2}$ se puede tomar como aproximación $\hat{\rho} = \frac{\bar{y}}{\bar{x}}$.

Este problema se ha estudiado por diversos autores como Tsui y Tang (2005), Ruben (2002), Bin Dong (2004) y la solución que se proporcionó ha recibido muchas críticas debido al uso de aproximaciones.

El problema fundamental es que no siempre es fácil proponer una cantidad pivotal, para de ahí calcular el intervalo de confianza, por lo que se hace necesario encontrar alternativas.

1.4.3. Uso explícito de la estadística suficiente minimal

Considérese la función de verosimilitud, función de θ para $x_1, \dots, x_n = \underline{x}$ fija, y puesto que $L(\theta; \underline{x})$, se puede factorizar para S suficiente. Esa factorización es el producto de dos funciones tales que:

$$L(\theta, \underline{x}) = \prod_{i=1}^n f(x_i; \theta) = g(s(\underline{x}); \theta) H(\underline{x}) \quad \forall \theta \in \Theta \quad (8)$$

donde g depende de la estadística suficiente y de θ , mientras que H sólo depende de la muestra observada.

Se puede concluir entonces que la verosimilitud es proporcional a la función de densidad de la estadística suficiente.

$$L(\theta; \underline{x}) \propto g(s(\underline{x}); \theta) \quad (9)$$

Considérese elegir $S(\underline{x})$ minimal, y basar en ella la inferencia sobre θ , así se cumple automáticamente el principio de verosimilitud de Fisher.

Al aplicar técnica de cantidad pivotal para el cálculo del intervalo, pero usando una exacta, se busca una cantidad aleatoria que sea función de la muestra, a través de la estadística suficiente minimal y del parámetro en cuestión, cuya distribución sea única para cualquier $\theta \in \Theta$.

Para construir cantidades pivotaes exactas puede usarse un procedimiento que se ha llamado universal, que se basa en la transformación con la integral de probabilidad usando la función de distribución de la estadística suficiente minimal, siempre y cuando esta última sea continua.

Sea $G^*(s; \theta)$ la función de distribución continua de $S(\underline{X})$, a la que se le aplica la transformación con la integral de probabilidad, definiendo la cantidad pivotal exacta como:

$$Q(s(\underline{X}); \theta) = G^*(s(\underline{X}); \theta) \quad (10)$$

La cual tiene una distribución uniforme en el intervalo $(0,1)$, independiente de θ .

Ejemplo 1.6.

Supóngase que X_1, \dots, X_n son variables aleatorias iid $f(x, \theta)$, donde

$$f(x, \theta) = \begin{cases} \frac{1}{\theta} & \text{si } 0 \leq x \leq \theta \\ 0 & \text{en otro caso (eoc)} \end{cases}$$

La estadística suficiente para θ , al aplicar el teorema de factorización queda

$$S(\underline{X}) = \max\{X_1, \dots, X_n\}, \text{ por lo que } L(\theta; \underline{x}) = \frac{1}{\theta^n} \propto \frac{1}{\theta^n} I\left(\max_{1 \leq i \leq n} x_i \leq \theta\right).$$

La estadística $S(\underline{X})$ tiene función de distribución continua dada por:

$$\begin{aligned} G^*(s; \theta) &= P(S \leq s) \\ &= P(X_1 \leq s, \dots, X_n \leq s) \\ &= \prod_{i=1}^n P(X_i \leq s) \\ &= \prod_{i=1}^n F_X(s; \theta) \end{aligned}$$

Donde $F_X(s; \theta)$ es la función de distribución acumulada de las variables X_1, \dots, X_n ,

$$F_X(x; \theta) = \begin{cases} 0 & \text{si } x \leq 0 \\ \frac{x}{\theta} & \text{si } 0 < x < \theta \\ 1 & \text{si } x \geq \theta \end{cases}$$

Calculando con esto la cantidad pivotal exacta se tiene,

$$Q(S(\underline{X}); \theta) = G^*(S(\underline{X}); \theta) = \left(\frac{S}{\theta}\right)^n \sim U(0,1)$$

El intervalo de confianza para θ al $(1-\alpha) \times 100\%$ quedaría $\left(\frac{S}{q_2^{1/n}}, \frac{S}{q_1^{1/n}}\right)$ con q_1 y q_2

cualquier par de cuantiles de una función de distribución acumulada de la distribución Uniforme $(0, 1)$, que satisfagan (5).

Si se quiere minimizar el intervalo, entonces: $LI(x_1, \dots, x_n) = \frac{S}{q_1^{1/n} - q_2^{1/n}} = S \left(\frac{1}{q_1^{1/n}} - \frac{1}{q_2^{1/n}} \right)$,

se esperaría escoger q_1 y q_2 , dado que es de la $U(0,1)$, tales que $q_1 = k\alpha$ y por tanto para cumplir la confianza de $(1-\alpha) \times 100\%$ q_2 debe ser $1 - (1-k)\alpha$, donde $k \in (0,1)$,

entonces $LI(x_1, \dots, x_n) = S \left(\frac{1}{(k\alpha)^{1/n}} - \frac{1}{(1-(1-k)\alpha)^{1/n}} \right)$, como S es fija, entonces se

considera a $R(k) = \frac{1}{(k\alpha)^{1/n}} - \frac{1}{(1-(1-k)\alpha)^{1/n}}$ la función a minimizar, tomando su

derivada, se tiene:

$$\begin{aligned} \frac{\partial R(k)}{\partial k} &= -\frac{\alpha}{n} (k\alpha)^{-\frac{1}{n}-1} + \frac{\alpha}{n} (1-(1-k)\alpha)^{-\frac{1}{n}-1} \\ &= \frac{\alpha^{-\frac{1}{n}}}{n} \left[\left(\frac{1-\alpha}{\alpha} + k \right)^{-\frac{1+n}{n}} - k^{-\frac{1+n}{n}} \right] \end{aligned}$$

La función $R(k)$ es monótona decreciente en el intervalo $[0,1]$, por lo que alcanza el mínimo valor cuando $k=1$; al sustituir este valor en $q_1 = k\alpha$ y en $q_2 = 1 - (1-k)\alpha$, resulta $q_1 = \alpha$ y $q_2 = 1$; por lo que el intervalo más pequeño al $(1-\alpha) \times 100\%$ de confianza para θ es de la forma $\left(S, \frac{S}{\alpha^{1/n}} \right)$.

Ejemplo 1.7.

Sean X_1, \dots, X_n v. a. iid Rayleigh(θ), es decir $f(x, \theta) = \frac{x}{\theta} e^{-\frac{x^2}{2\theta}}; x > 0$, la estadística suficiente minimal para θ es $S(\underline{X}) = \sum_{i=1}^n X_i^2$, cuya función de densidad de probabilidad es gama ($\alpha = n, \beta = 2\theta$) ⁶ y función de distribución continua:

$$G^*(s; \theta) = \begin{cases} 0 & \text{si } s \leq 0 \\ \int_0^{\frac{s}{2\theta}} \frac{w^{n-1} e^{-w}}{\Gamma(n)} dw & \text{si } 0 < s < \infty \end{cases}$$

Se tiene que la cantidad pivotal exacta es:

$$Q(s; \theta) = G^*(s; \theta) = \frac{1}{\Gamma(n)} \int_0^{\frac{s}{2\theta}} w^{n-1} e^{-w} dw$$

con distribución uniforme (0,1).

Para mostrar el cálculo del intervalo de confianza se hará la suposición de que $n = 4$, por lo que este intervalo para θ al $(1-\alpha) \times 100\%$ es:

$$P(q_1 < Q < q_2) = 1 - \alpha$$

$$P\left(q_1 < -\frac{z^3 e^{-z/2\theta}}{48\theta^3} - \frac{z^2 e^{-z/2\theta}}{8\theta^2} - \frac{z e^{-z/2\theta}}{2\theta} + e^{-z/2\theta} + 1 < q_2\right) = 1 - \alpha$$

con q_1 y q_2 cualquier par de cuantiles de una función de distribución acumulada de la Uniforme (0, 1), que satisfagan (5).

Es importante hacer notar que efectivamente se puede encontrar una expresión para la cantidad pivotal exacta, pero ésta no siempre es sencilla, como se aprecia en el ejemplo

⁶ $g^*(s; \theta) = \frac{\left(\frac{1}{2\theta}\right)^n s^{n-1} e^{-\frac{s}{2\theta}}}{\Gamma(n)}; s > 0$

1.7. y esto se convierte en la principal limitante para encontrar el intervalo de confianza para θ , ya que el despeje del parámetro no resulta fácil.

1.4.4. Importancia de los valores q_1 y q_2

Se pueden escoger muchos pares de valores q_1 y q_2 para cualquier nivel fijo $\gamma = 1 - \alpha$ que satisfaga $P(q_1 < Q < q_2) = \gamma$, pero se busca que el intervalo de confianza para θ , sea lo más pequeño posible; esto se logra escogiendo aquéllos que hagan la longitud del intervalo para θ lo más pequeña posible en promedio. Si esta longitud no es aleatoria, se pueden escoger cantidades q_1 y q_2 tales que $g(q_2) = g(q_1)$, donde $g(\cdot)$ es la distribución de la cantidad pivotal, con lo cual se asegura que $q_2 - q_1$ sea mínima para $\gamma = 1 - \alpha$ fija. Si la longitud es aleatoria, entonces se puede recurrir al intervalo más corto en promedio, o sea, con longitud esperada más pequeña, pero ese no es el interés.

1.4.5. Aplicación del procedimiento pivotal con uso de la cantidad pivotal asintótica

Como fue referido anteriormente no siempre es posible encontrar la cantidad pivotal y su distribución exacta, en estos casos se utiliza una cantidad pivotal aproximada para calcular intervalos de confianza aproximados para θ . Una forma clásica de enfrentar el problema, es usando el estimador $\hat{\theta}$ (máximo verosímil), el cual se distribuye aproximadamente normal con media θ y varianza $Var(\theta)$, así

$$Q(\theta) = \frac{\hat{\theta} - \theta}{\sqrt{Var(\theta)}} \sim N(0,1)$$

$Q(\theta)$ es llamada la cantidad pivotal asintótica para θ .

Cuando se tiene que $Var(\theta)$ es desconocida se puede sustituir por su estimador $\sqrt{Var(\hat{\theta})}$ en la cantidad pivotal y por la simetría de la función normal se tendría que $\Phi(q) - \Phi(-q) = p$, con $\Phi(x)$ la función de distribución acumulada normal, entonces,

$$P\left(-q \leq \frac{\hat{\theta} - \theta}{\sqrt{Var(\hat{\theta})}} \leq q\right) \approx \gamma \quad (11)$$

el cual tiene una confianza aproximada de $(1 - \alpha) \times 100\%$.

Cabe observar que esto es la famosa Regla Empírica mostrada en la sección 1.2.1.

Ejemplo 1.8.

Del ejemplo 1.2. se encontró que de la función de densidad dada por $f(x; \theta) = \frac{3x^2}{\theta^3}$; $0 \leq x \leq \theta$, con estimador máximo verosímil de θ $Y_n = \max\{X_1, \dots, X_n\}$ y con n grande, se tuvo que la cantidad pivotal asintótica fue:

$$Q_n = \frac{(3n+1)Y_n - 3n\theta}{\theta \sqrt{\frac{3n}{3n+2}}} \sim N(0,1)$$

Y calculando el intervalo de confianza al $(1 - \alpha) \times 100\%$ para θ quedó:

$$\left(\frac{y_n(3n+1)\sqrt{1+\frac{2}{3n}}}{q + \sqrt{3n(3n+2)}}, \frac{y_n(3n+1)\sqrt{1+\frac{2}{3n}}}{-q + \sqrt{3n(3n+2)}} \right), \text{ donde } q \text{ es el cuantil } 1 - \frac{\alpha}{2} \text{ de } \Phi(x).$$

Otro procedimiento estudiado por Sprott para la construcción de cantidades pivotaes asintóticas es usando exclusivamente la función de verosimilitud $L(\theta, \underline{x}) = \prod_{i=1}^n f(x_i; \theta)$ y

las cualidades vistas en la sección 1.1. Al observar que $E\left(\frac{\partial \log L(\theta; \underline{x})}{\partial \theta}\right) = 0$ y que

$$\text{Var}\left(\frac{\partial \log L(\theta; \underline{x})}{\partial \theta}\right) = -E\left(\frac{\partial^2 \log L(\theta; \underline{x})}{\partial \theta^2}\right)$$

Se tiene la cantidad pivotal:

$$Q(\theta) = \frac{\frac{\partial \log L(\theta; \underline{x})}{\partial \theta}}{\sqrt{E\left(\frac{\partial \log L(\theta; \underline{x})}{\partial \theta}\right)^2}}$$

$$Q(\theta) = \frac{\frac{\partial \log L(\theta; \underline{x})}{\partial \theta}}{\sqrt{-E\left(\frac{\partial^2 \log f(\theta; \underline{x})}{\partial \theta^2}\right)}} \quad (12)$$

donde $-E\left(\frac{\partial^2 \log f(\theta; \underline{x})}{\partial \theta^2}\right) = I_E(\theta)$ es la cantidad de información esperada de Fisher para la muestra X_1, \dots, X_n . El intervalo aproximado para θ al $(1 - \alpha) \times 100\%$ será:

$$P\left(-q < \frac{\frac{\partial \log L(\theta; \underline{x})}{\partial \theta}}{\sqrt{I_E(\theta)}} < q\right) = 1 - \alpha \quad (13)$$

donde q será el cuantil $\left(1 - \frac{\alpha}{2}\right) \times 100\%$ de una $N(0,1)$ y los intervalos obtenidos como solución en θ de

$$-q < \frac{\frac{\partial \log L(\theta; \underline{x})}{\partial \theta}}{\sqrt{I_E(\theta)}} < q$$

contendrán el verdadero valor de θ con una frecuencia relativa de $1-\alpha$ aproximadamente. El problema de esto es que es cierto para muestras grandes y depende de qué tan rápido converge $Q(\theta)$ a la normal estándar y por lo tanto esto está condicionado a la "normalidad" de la función de verosimilitud de θ .

Este problema lo enfrentó y revisó Sprott (1973, 1984) y encontró que cuando la función de verosimilitud es muy asimétrica o bien tiene colas pesadas, estimaciones del tipo $\theta = \hat{\theta} \pm \sigma$ son engañosas.

Para construir cantidades pivótales asintóticas, Sprott (1975) se basó, como ya se mencionó, en la estimación máximo verosímil. Este planteamiento considera las propiedades frecuenciales de la estimación de θ , la cual es solución de la ecuación

$$\begin{aligned} T(\underline{x}; \theta) &= \left. \frac{\partial \log L(\theta; \underline{x})}{\partial \theta} \right|_{\theta=\hat{\theta}} \\ &= \left. \frac{\partial \log f(\underline{x}; \theta)}{\partial \theta} \right|_{\theta=\hat{\theta}} \\ &= 0 \end{aligned} \tag{14}$$

Como se indicó en la sección 1.1.3 $T(\underline{X}; \theta)$ es la función score o puntaje.

Bajo condiciones de regularidad apropiadas, se tiene:

$$E_{\theta}(T(\underline{X}; \theta)) = 0 \tag{15}$$

y

$$Var_{\theta}(T(\underline{X}; \theta)) = -E_{\theta} \left(\frac{\partial^2 \log(\underline{X}; \theta)}{\partial \theta^2} \right) = I_E(\theta) \tag{16}$$

Donde $I_E(\theta)$ es la información esperada de Fisher para la muestra, la cual es igual a $ni_E(\theta)$, que es la información esperada de Fisher por unidad muestral x , con

$$ni_E(\theta) = -E_\theta \left(\frac{\partial^2 \log(X_i; \theta)}{\partial \theta^2} \right).$$

Como los resultados son exactos con muestras de cualquier tamaño, $T(\underline{X}; \theta)$ es estable en la media.

Por otra parte el score puede expresarse como:

$$\frac{\partial \log \prod_{i=1}^n f(X_i; \theta)}{\partial \theta} = \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f(X_i; \theta) \quad (17)$$

y aplicando el Teorema Central del Límite, se tiene que asintóticamente $\frac{1}{\sqrt{n}}T(\underline{X}; \theta)$ tiene una distribución límite, $N(0, ni_E(\theta))$.

El procedimiento de máxima verosimilitud ha sido usado para producir estimadores puntuales $\hat{\theta}$, que son asintóticamente insesgados, es decir, $E_\theta(\hat{\theta}) \approx \theta$ para n grande y $Var_\theta(\hat{\theta}) \approx I_E^{-1}(\theta)$; para n grande. Así:

$$\hat{\theta} \sim N(0, I_E^{-1}(\theta)). \quad (18)$$

Como los resultados son asintóticos, la cantidad

$$\hat{\theta} - \theta \quad (19)$$

puede verse como la discrepancia en la estimación y es asintóticamente estable en la media.

En el trabajo desarrollado por Sprott (1975) aparece una lista de cantidades pivotaes asintóticas, por ejemplo:

$$T_1 = \frac{T}{\sqrt{\text{Var}_\theta(T_\theta(\underline{X}; \theta))}} \quad (20)$$

$$T_1 = \frac{T}{\sqrt{I_E(\theta)}} \quad \simeq N(0,1)$$

Tomando (19) y sabiendo (18)

$$Q_1 = (\hat{\theta} - \theta) \sqrt{I_E(\theta)} \simeq N(0,1) \quad (21)$$

Tanto T_1 como Q_1 usan la varianza exacta dada en (16).

Cuando no se tiene $I_E(\theta)$ se usan estimadores de éste que son:

$$I_E(\hat{\theta}) = -E_\theta \left(\frac{\partial^2 \log f(\underline{X}; \theta)}{\partial \theta^2} \right) \Bigg|_{\theta=\hat{\theta}} \quad (22)$$

$$I_o(\hat{\theta}) = - \frac{\partial^2 \log f(\underline{x}; \hat{\theta})}{\partial \theta^2} \Bigg|_{\theta=\hat{\theta}} \quad (23)$$

Así

$$T_2 = \frac{T}{\sqrt{I_E(\hat{\theta})}} \simeq N(0,1) \quad (24)$$

$$T_3 = \frac{T}{\sqrt{I_o(\hat{\theta})}} \simeq N(0,1) \quad (25)$$

y en el caso de (19) se tiene

$$Q_2 = \frac{\hat{\theta} - \theta}{\sqrt{I_E(\hat{\theta})}} \sim N(0,1) \quad (26)$$

utilizando I_o queda:

$$Q_3 = (\hat{\theta} - \theta) \sqrt{I_o(\hat{\theta})} \sim N(0,1) \quad (27)$$

Bajo las condiciones apropiadas, todas las cantidades pivotaes asintóticas T_1, T_2, T_3, Q_1, Q_2 y Q_3 ; son asintóticamente equivalentes, es decir, para valores grandes de n , todas tienen la distribución límite $N(0,1)$; por lo que se puede calcular el intervalo de confianza y queda:

$$p(-b \leq Q_n(\underline{X}; \theta) \leq b) \doteq 1 - \alpha \quad (28)$$

con b cuantil $1 - \frac{\alpha}{2}$ de $N(0,1)$.

Los intervalos obtenidos como solución en θ de $-b \leq Q_n(\underline{X}; \theta) \leq b$ contendrán el verdadero valor de θ con una frecuencia relativa de $1 - \alpha$ aproximadamente.

Ejemplo 1.9.

Sea X_1, \dots, X_n v. a. iid Bernoulli (p) ⁷, con $L(p; \underline{x}) = p^{\sum x_i} (1-p)^{n-\sum x_i}$

⁷ $f(x; p) = p^x (1-p)^{1-x}$; $x = 0, 1$

$$\frac{\partial}{\partial p} \ln L(p; \underline{x}) = \frac{\sum x_i}{p} - \frac{n - \sum x_i}{1-p}$$

$$-E \left(\frac{\partial^2 \ln L(p; \underline{x})}{\partial p^2} \right) = \frac{n}{p} + \frac{n}{1-p}.$$

Por lo que la cantidad pivotal se puede construir como:

$$Q(p) = \frac{\frac{n\bar{x}}{p} - \frac{n - n\bar{x}}{1-p}}{\sqrt{\frac{n}{p} + \frac{n}{1-p}}} = \frac{\bar{x} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1).$$

Esta cantidad, tiene el problema de que el parámetro p se encuentra tanto en el subradical como en el numerador y su despeje no es inmediato, por lo que usando una de las cantidades pivotaes propuestas por Sprott queda:

$$Q_2 = \frac{\frac{\partial}{\partial \theta} \ln L(p; \underline{x})}{I_{E(\hat{\theta})}} \sim N(0,1)$$

por lo que la nueva cantidad pivotal es:

$$Q_2(p) = \frac{\bar{x} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \sim N(0,1)$$

donde $\hat{p} = \frac{\sum x_i}{n}$ es el estimador máximo verosímil de p .

El intervalo aproximado de confianza al $(1 - \alpha) \times 100\%$ es:

$$\left(\bar{x} - q\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \bar{x} + q\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

donde q es el cuantil $1 - \frac{\alpha}{2}$ de una distribución $N(0,1)$.

Se hace la observación de que este procedimiento basado en un acercamiento asintótico proporcionará resultados "razonables" en la medida en que la cantidad pivotal Q_2 , en efecto tenga una "buena" aproximación normal. Si se tuviera un caso extremo considerando una muestra de tamaño 10 ensayos Bernoulli, sólo se observara un éxito, jamás se debiera usar este procedimiento, ya que no se tendrán resultados razonables.

Todo lo visto en la construcción de cantidades pivotaes muestra cómo el procedimiento basado en ellas, consiste en poder encontrar una desigualdad doble en la que el parámetro θ quede contenido, lo cual no siempre es fácil y es necesario tener claridad en la manipulación que permitan esas cantidades.

Capítulo 2

Argumento fiducial y enfoque Bayesiano

Introducción

En este capítulo se revisará, a través de los ejemplos del capítulo anterior, la aplicación del procedimiento debido a Fisher, que se basa en el controvertido “argumento fiducial”, obteniéndose intervalos fiduciales, cuya interpretación es diferente a la de los intervalos de confianza vistos anteriormente. También se encontrarán los llamados intervalos de “más alta densidad final”, para cada uno de los ejemplos que se verán en el argumento fiducial, considerando diferencias y similitudes. Se revisa, también, la obtención de las distribuciones finales o *a-posteriori* y las distribuciones fiduciales en el caso en que las iniciales (o *a-priori*), son elegidas como no-informativas.

2.1. Argumento fiducial

En 1930 Fisher introduce un procedimiento, conocido como *fiducial*, para hacer inferencias sobre el valor de un parámetro θ . Este modo lo hace como alternativo al procedimiento Bayesiano, en que se obtiene una distribución sobre θ a partir de una distribución inicial y la muestra (a través de la estadística suficiente minimal). Este argumento fiducial no utiliza una distribución inicial, se basa tan sólo en las observaciones disponibles (la muestra), o sea (x_1, x_2, \dots, x_n) .

En este planteamiento θ juega el papel lógico de una variable aleatoria, *sin serlo* y S , la estadística suficiente minimal, "hace el papel" de un parámetro que define la *distribución fiducial* de θ .

Se supone que la muestra, $\underline{x} = (x_1, x_2, \dots, x_n)$, ya ha sido observada y con ella la estadística suficiente $S = s_0$ ha quedado determinada. Suponiendo la existencia de una cantidad pivotal exacta $Q(s, \theta)$ creciente en θ , para cada s fija, entonces se cumple que $Q(s_0, \theta) \leq q \Leftrightarrow \theta \leq Q^{-1}(s_0, q)$, por lo que, a la desigualdad:

$$\theta \leq Q^{-1}(s_0, q), \quad (1)$$

se le asocia una "*probabilidad fiducial*"

$$p_{fid}(\theta \leq Q^{-1}(s_0, q)) = G(q) \quad (2)$$

con G la función de distribución de Q .

A $Q^{-1}(s_0, q)$ se le llama θ^* y es claro que depende de q y de s_0 , el valor que tomó la estadística suficiente minimal S . Puede notarse en:

$$\theta^* = Q^{-1}(s_0, q) \Leftrightarrow Q(s_0, \theta^*) = q. \quad (3)$$

La función de distribución fiducial surge reescribiendo (2), entonces

$$F_{fid}(\theta^*) = p_{fid}(\theta \leq \theta^*) = G(Q(s_0, \theta^*)) = \gamma \quad (4)$$

Para obtener la función de densidad fiducial, en el caso continuo, se debe suponer que G tiene derivadas continuas y $Q(s_0, \cdot)$ es diferenciable respecto a θ , entonces:

$$\begin{aligned}
f_{fid}(\theta^*) &= \frac{dF_{fid}(\theta^*)}{d\theta^*} \\
&= \frac{dG(Q(s_0, \theta^*))}{d\theta^*} \\
&= g(Q(s_0, \theta^*)) \cdot \frac{\partial Q(s_0, \theta^*)}{\partial \theta^*}
\end{aligned} \tag{5}$$

donde g es la función de densidad de la cantidad pivotal Q .

Se puede tener el caso de que la cantidad pivotal Q sea una función monótona decreciente en θ para cada s , entonces (4) quedaría:

$$\begin{aligned}
F_{fid}(\theta^*) &= 1 - p_{fid}(\theta \leq \theta^*) \\
&= 1 - G(Q(s_0, \theta^*)) \\
&= \gamma
\end{aligned} \tag{6}$$

y la función de densidad fiducial sería

$$f_{fid}(\theta^*) = -g(Q(s_0, \theta^*)) \cdot \frac{\partial Q(s_0, \theta^*)}{\partial \theta^*} \tag{7}$$

Considerando los dos casos se tiene que la función de densidad fiducial para $Q(s; \theta)$ monótona en θ para cada s_0 es:

$$f_{fid}(\theta^*) = g(Q(s_0, \theta^*)) \cdot \left| \frac{\partial Q(s_0, \theta^*)}{\partial \theta^*} \right| \tag{8}$$

con θ^* relacionada con γ y s_0 a través de

$$\gamma = G(Q(s_0, \theta^*)) \quad (9)$$

La función de distribución de θ , se escribe como:

$$F_{fid}(\theta^*) = \begin{cases} G(Q(s_0, \theta^*)) & ; \quad Q \text{ es creciente en } \theta \\ 1 - G(Q(s_0, \theta^*)) & ; \quad Q \text{ es decreciente en } \theta \end{cases} \quad (10)$$

Para obtener el intervalo fiducial central para θ , con "probabilidad fiducial" γ y basado en el valor observado $S = s_0$ (como parámetro), hay que encontrar cantidades θ_1^* y θ_2^* tales que

$$p_{fid}(\theta_1^* \leq \theta \leq \theta_2^*) = F_{fid}(\theta_2^*) - F_{fid}(\theta_1^*) = \gamma \quad (11)$$

Si además se quiere que ese intervalo sea el más pequeño, se debe satisfacer, que θ_1^* y θ_2^* cumplan:

$$f_{fid}(\theta_1^*) = f_{fid}(\theta_2^*) \quad (12)$$

La interpretación de los intervalos fiduciales no es la misma que la de los intervalos de confianza, así en este caso, se dice *que para una muestra particular que se observó y que produjo el valor s_0 , el valor del parámetro que es desconocido, θ , se encontrará entre θ_1^* y θ_2^* con una "probabilidad fiducial" γ .*

Ejemplo 2.1

Tomando el ejemplo 1.1 se calculará su intervalo fiducial. Sea X_1, \dots, X_n una muestra aleatoria de una $Exp(\lambda)$, $f(x; \lambda) = \lambda e^{-\lambda x} I_{(0, \infty)}(x)$, la estadística suficiente minimal es

$S(\underline{X}) = \sum_{i=1}^n X_i$ donde $S(\underline{X}) \sim \Gamma\left(\frac{1}{\lambda}, n\right)$. En el ejemplo 1.1 se dio la cantidad pivotal, la

cual es:

$$Q(S; \lambda) = \lambda S \sim \Gamma(n, 1)$$

Se tiene entonces

$$G(q) = \int_0^{\lambda s} \frac{z^{n-1}}{\Gamma(n)} e^{-z} dz$$

Aplicando (10) se puede calcular la distribución fiducial, la cual es:

$$F_{fid}(\lambda^*) = \int_0^{\lambda^* s_0} \frac{z^{n-1}}{\Gamma(n)} e^{-z} dz$$

Aplicado (5) se tiene que la función de densidad fiducial es:

$$f_{fid}(\lambda^*) = \frac{1}{\Gamma(n)} (\lambda^* s_0)^{n-1} s_0 e^{-(\lambda^* s_0)}$$

Indicando que dado el valor observado $S = s_0$, λ tiene como distribución fiducial una

gama $\left(n, \frac{1}{s_0}\right)$.

El intervalo fiducial sería calculado construyendo λ_1^* y λ_2^* siguiendo (11):

$$\int_{\lambda_1^* s_0}^{\lambda_2^* s_0} \frac{z^{n-1}}{\Gamma(n)} e^{-z} dz = \gamma$$

Si además se requiere que sea mínimo, entonces:

$$(\lambda_1^*)^{n-1} e^{-\lambda_1^* s_0} = (\lambda_2^*)^{n-1} e^{-\lambda_2^* s_0}$$

Resultando igual que al calculado en el ejemplo 1.3.

Existen ejemplos en los que no existe una cantidad pivotal clara, en el sentido clásico, de localización y escala; en general de modelo de grupo, en que este procedimiento puede ser utilizado, como se verá en el capítulo 3 con la distribución gama (α) y con la exponencial truncada.

2.1.1. Argumento fiducial cuando es explícita la función de distribución de la estadística suficiente minimal

Si la propiedad pedida a la pivotal $Q(s; \theta)$ resulta ser ⁸ la misma que la de la distribución $G_0(s; \theta)$, es decir, que sea continua en s y $G_0(s; \theta)$ función creciente en θ para cada s fija, entonces se cumple que:

$$Q(s, \theta) \leq \gamma \Leftrightarrow G_0(s; \theta) \leq \gamma \Leftrightarrow \theta \leq G_0^{-1}(s; \gamma) \quad (13)$$

(13) se convierte en:

$$\theta \leq G_0^{-1}(s_0, \gamma) \quad (14)$$

Donde G_0 es la función de distribución de S con probabilidad fiducial asociada:

⁸ Como $Q(s; \theta) = G_0(s; \theta) \sim U(0, 1)$, $p(Q(s; \theta) \leq q; \theta) = p(G_0(s; \theta) \leq q; \theta) = q, \quad \forall \theta \in \Theta$.

$$p_{fid}(\theta \leq G_0^{-1}(s_0, \gamma)) = \gamma \quad (15)$$

$$\theta^* = G_0^{-1}(s_0, \gamma) \Leftrightarrow G_0(s_0, \theta^*) = \gamma \quad (16)$$

Es importante señalar que θ^* está relacionada con γ por una $U(0,1)$ y con s_0 .

Con (15) y (16), se define la función de distribución fiducial:

$$\begin{aligned} F_{fid}(\theta^*) &= p_{fid}(\theta \leq \theta^*) \\ &= G_0(s_0, \theta^*) \\ &= \gamma. \end{aligned} \quad (17)$$

Para obtener a la función de densidad fiducial, se necesita que $G_0(s, \cdot)$ sea diferenciable con respecto a θ , para cada s entonces,

$$\begin{aligned} f_{fid}(\theta^*) &= \frac{dF_{fid}(\theta^*)}{d\theta^*} \\ &= \frac{\partial G_0(s_0, \theta^*)}{\partial \theta^*} \end{aligned} \quad (18)$$

Si se diera el caso de que $G_0(s, \theta)$ fuera continua y monótona decreciente en θ para cada s , entonces (17) se reescribe como:

$$\begin{aligned} F_{fid}(\theta^*) &= 1 - p_{fid}(\theta \leq \theta^*) \\ &= 1 - G_0(s_0, \theta^*) \\ &= \gamma \end{aligned} \quad (19)$$

y la función la densidad fiducial sería:

$$\begin{aligned}
f_{fid}(\theta^*) &= \frac{dF_{fid}(\theta^*)}{d\theta^*} \\
&= -\frac{\partial G_0(s_0, \theta^*)}{\partial \theta^*}
\end{aligned}
\tag{20}$$

Tomando (18) y (20) se escribe la densidad como

$$f_{fid}(\theta^*) = \left| \frac{\partial G_0(s_0, \theta^*)}{\partial \theta^*} \right|
\tag{21}$$

Y la función de distribución fiducial queda:

$$F_{fid}(\theta^*) = \begin{cases} G_0(s_0, \theta^*) & \text{si } G_0 \text{ creciente en } \theta \\ 1 - G_0(s_0, \theta^*) & \text{si } G_0 \text{ decreciente en } \theta \end{cases}
\tag{22}$$

Para el caso en que $G_0(s; \theta)$ corresponda a una distribución de S discreta, puede emularse el procedimiento descrito para el caso continuo, pero podría utilizarse tanto la versión de $G_0(s; \theta)$ continua por la derecha como continua por la izquierda (y todas las combinaciones convexas de ellas).

Sin entrar en detalles que limitan la obtención de una fiducial en el caso discreto (falta de unicidad como se señaló) y el que los intervalos resultantes no puedan hacerse para una γ arbitraria (por lo discreto de G_0), más adelante se mostrará un ejemplo en que al menos se propone una única fiducial utilizando un requerimiento de invarianza.

Ejemplo 2.2.

Considerando el ejemplo 2.1, la cantidad pivotal $Q(S, \lambda) = \lambda S$, con $S = \sum_{i=1}^n X_i$.

Tomando $Q(S, \lambda) = G_0(s; \lambda) \sim U(0, 1)$ con $G_0(s; \lambda) \sim \Gamma\left(\frac{1}{\lambda}, n\right)$.

Ya que se observó la muestra, se tiene que $S = s_o$, por lo que la función de distribución fiducial dada por (17) es:

$$F_{fid}(\lambda^*) = P(\lambda \leq \lambda^*) = \int_0^{s_o} \frac{1}{\Gamma(n)} (\lambda^*)^n z^{n-1} e^{-\lambda^* z} dz$$

y por lo tanto la densidad fiducial quedaría como:

$$\begin{aligned} f_{fid}(\lambda^*) &= \frac{\partial G_0(s_o, \lambda^*)}{\partial \lambda^*} \\ &= \frac{\partial}{\partial \lambda} \int_0^{s_o} \frac{1}{\Gamma(n)} (\lambda^*)^n z^{n-1} e^{-\lambda^* z} dz \end{aligned}$$

Haciendo un cambio de variable, donde $w = \lambda^* z$; $z = \frac{w}{\lambda^*}$; $dw = \lambda^* dz$, entonces

$$\begin{aligned} f_{fid}(\lambda^*) &= \frac{\partial}{\partial \lambda} \int_0^{\lambda^* s_o} \frac{1}{\Gamma(n)} (\lambda^*)^n \left(\frac{w}{\lambda^*}\right)^{n-1} e^{-w} \frac{1}{\lambda^*} dw \\ &= \frac{\partial}{\partial \lambda} \frac{1}{\Gamma(n)} \int_0^{\lambda^* s_o} w^{n-1} e^{-w} dw \\ &= \frac{1}{\Gamma(n)} (\lambda^* s_o)^{n-1} e^{-\lambda^* s_o} \sim \Gamma\left(\alpha = n, \beta = \frac{1}{\lambda^*}\right) \end{aligned}$$

Que es la misma que se exhibió en el ejemplo 2.1.

Ejemplo 2.3

Tomando el ejemplo 1.6 donde X_1, \dots, X_n son v.a. iid $f(x; \theta)$ con

$$f(x; \theta) = \begin{cases} \frac{1}{\theta} & 0 \leq x \leq \theta; \quad \theta > 0 \\ 0 & \text{eoc} \end{cases}$$

El intervalo de confianza al $(1 - \alpha) \times 100\%$ resultó $\left(\frac{S}{q_2^{1/n}}, \frac{S}{q_1^{1/n}} \right)$ y el más pequeño fue

$\left(S, \frac{S}{\alpha^{1/n}} \right)$, utilizando ahora lo visto en 2.1.2, se toma primero a la estadística suficiente

minimal $S = \text{máx}\{X_1, \dots, X_n\}$ con

$$G_0(s; \theta) = \begin{cases} 0 & \text{si } s < 0 \\ \left(\frac{s}{\theta} \right)^n & \text{si } 0 \leq s \leq \theta \\ 1 & \text{si } \theta < s < \infty \end{cases}$$

derivando esta función se obtiene $g_0(s; \theta)$

$$g_0(s; \theta) = \begin{cases} \frac{n}{\theta} \left(\frac{s}{\theta} \right)^{n-1} & \text{si } 0 \leq s \leq \theta \\ 0 & \text{eoc} \end{cases}$$

y la cantidad pivotal queda:

$$\begin{aligned} Q(S, \theta) &= G_0(S; \theta) \\ &= \left(\frac{S}{\theta} \right)^n \sim U(0, 1). \end{aligned}$$

Cuando ya se observó la muestra, la función de distribución fiducial queda:

$$\begin{aligned}
F_{fid}(\theta^*) &= 1 - G_0(s_0; \theta^*) \\
&= \begin{cases} 1 & \text{si } s_0 < 0 \\ 1 - \left(\frac{s_0}{\theta^*}\right)^n & \text{si } 0 \leq s_0 \leq \theta^* \\ 0 & \text{si } \theta^* < s_0 < \infty \end{cases} \\
&= \begin{cases} 1 - \left(\frac{s_0}{\theta^*}\right)^n & \text{si } s_0 \leq \theta^* \\ 0 & \text{si } \theta^* < s_0 \end{cases}
\end{aligned}$$

Derivando esta última función (aplicando 19) se obtiene:

$$\begin{aligned}
f_{fid}(\theta^*) &= -\frac{\partial G_0(s_0; \theta^*)}{\partial \theta^*} \\
f_{fid}(\theta^*) &= \begin{cases} n \left(\frac{s_0}{\theta^*}\right)^{n-1} \left(\frac{s_0}{(\theta^*)^2}\right) & \text{si } s_0 \leq \theta^* \\ 0 & \text{si } \theta^* < s_0 \end{cases} \\
&= \begin{cases} \frac{ns_0^n}{(\theta^*)^{n+1}} & \text{si } s_0 \leq \theta^* \\ 0 & \text{si } \theta^* < s_0 \end{cases}
\end{aligned}$$

Para obtener el intervalo fiducial, se hace:

$$p_{fid}(\theta \leq Q^{-1}(s_0, q_\gamma)) = G(q_\gamma) = \gamma$$

es decir,

$$\begin{aligned}
p_{fid}(\theta \leq \theta_2^*) &= F_{fid}(\theta_2^*) = \gamma \\
&= 1 - \left(\frac{s_0}{\theta_2^*}\right)^n = \gamma
\end{aligned}$$

Quedando la ecuación

$$\left(\frac{S_0}{\theta_2^*}\right)^n = 1 - \gamma = \alpha$$

$$\frac{S_0}{\theta_2^*} = \alpha^{1/n}$$

$$\frac{S_0}{\alpha^{1/n}} = \theta_2^*$$

Dejando el mismo resultado visto en el ejemplo 1.6.

Cuando se ha obtenido la función de distribución fiducial de θ , con ella se puede construir la correspondiente función de distribución fiducial de cualquier función uno a uno de θ , $g(\theta)$, considerando a ambas como variables aleatorias.

En los ejemplos anteriores fue posible resumir toda la información proporcionada por la muestra sobre el parámetro θ en una estadística suficiente minimal S de dimensión igual al parámetro, pero no siempre es posible encontrar una estadística así, aunque puede suceder que dicha estadística esté en correspondencia uno a uno con una estadística unidimensional (R) otra estadística ancilaria⁹ (U). Estas dos estadísticas son tales que la función de verosimilitud se puede expresar como:

$$L(\theta; \underline{x}) = f(\underline{x}; \theta) = f(R, U; \theta) f(\underline{x} / R, U)$$

Donde la distribución condicional de \underline{x} dada R y U no depende de θ , por lo que

$$L(\theta; \underline{x}) \propto f(R, U; \theta) = f(R; \theta / U) f(U).$$

Como U es ancilaria, la verosimilitud es proporcional a

⁹ La estadística ancilaria indica que su distribución no depende de θ .

$$L(\theta; \underline{x}) \propto f(R; \theta/U)$$

Por esto se puede basar las inferencias sobre θ en la distribución condicional de R dado U , en este caso se dice en la literatura que R es una *estadística exhaustiva o condicionalmente suficiente*.

Si la distribución condicional $f(R; \theta/U)$, es continua en R y monótona en θ para cada valor de R y el valor fijo y observado de U , se puede entonces, encontrar la distribución fiducial de θ como en (22) si, además $f(R; \theta/U)$ es una función diferenciable con respecto θ para cada R y el valor fijo de U , entonces la densidad fiducial se puede encontrar como en (21) (Sprott, 1967).

2.1.2. Argumento fiducial implícito

Esta variante permite obtener intervalos fiduciales para θ sin hacer explícita la función de distribución fiducial de θ .

El procedimiento es el siguiente:

1. Sea S una estadística suficiente minimal (uni-dimensional), con $G_0(s; \theta)$ su función de distribución.

Para cada $\gamma = 1 - \alpha$, $\gamma \in (0, 1)$ $\exists s_\gamma(\theta)$ tal que

$$P(S \leq s_\gamma; \theta) = G_0(s_\gamma(\theta), \theta) = \gamma \quad (23)$$

2. Se construye $A(s_0) = \{\theta; s_0 \leq s_\gamma(\theta)\}$ (24)

conjunto de valores de θ , donde $G_0(s, \bullet)$ es creciente en θ para cada valor s y cuando $S = s_0$.

$$3. \text{ Considérese } \theta_1 \in A(s_0) \Rightarrow s_0 \leq s_\gamma(\theta_1) \Leftrightarrow G_0(s_0; \theta_1) \leq G_0(s_\gamma(\theta_1); \theta_1) = \gamma \quad (25)$$

$$4. \text{ Tómesese } \theta_2 < \theta_1 \Rightarrow G_0(s_0; \theta_2) \leq G_0(s_0; \theta_1) \quad (26)$$

por ser $G_0(s; \theta)$ creciente y aplicando (24) se concluye

$$G_0(s_0; \theta_2) \leq G_0(s_\gamma(\theta_2); \theta_2) = \gamma \quad (27)$$

$$\Rightarrow s_0 \leq s_\gamma(\theta_2) \therefore \theta_2 \in A(s_0).$$

5. Esto indica que $A(s_0)$ tiene la forma del intervalo $(-\infty, \theta^*]$, donde θ^* es tal que $s_0 = s_\gamma(\theta^*)$.

6. Al conjunto $A(s_0)$ se le conoce como el "*intervalo fiducial*" de θ de "*probabilidad fiducial*" γ , para $S = s_0$, es decir, $p_{fid}(\theta \in (-\infty, \theta^*]) = \gamma$. (28)

Para encontrar la función de distribución fiducial de θ se retoma que:

Para $\theta \in A(s_0) \Leftrightarrow \theta \in (-\infty, \theta^*] \Leftrightarrow s_0 \in (-\infty, s_\gamma(\theta^*)]$, con s_0 fija, $\gamma \in (0,1)$; por lo que para cada valor de γ , $\exists \theta^*(\gamma)$ que satisface (27). Como al observar γ como función de θ^* a través de $s_0 = s_\gamma(\theta^*)$, entonces:

$$p_{fid}(\theta \in (-\infty, \theta^*]) = \gamma(\theta^*) \quad (29)$$

Al variar θ^* se tendrá que

$$F_{fid}(\theta^*) = \gamma(\theta^*) \quad (30)$$

Si se tuviera el caso de que $G_0(s, \theta)$ fuera decreciente en θ , para cada s , el conjunto $A(s_0)$ de θ daría intervalos frecuenciales para θ del tipo $[\theta^*, \infty)$.

Ejemplo 2.4.

Regresando al ejemplo 2.3. con X_1, \dots, X_n v.a. iid $U(0, \theta)$ y con $S(\underline{X}) = \text{máx}\{X_1, \dots, X_n\}$, estadística suficiente minimal cuyas funciones de densidad y distribución son:

$$g_0(s, \theta) = \begin{cases} \frac{n}{\theta^n} s^{n-1} & \text{si } 0 < s \leq \theta \\ 0 & \text{eoc} \end{cases} \quad \text{y } G_0(s, \theta) = \begin{cases} 0 & \text{si } s < 0 \\ \frac{s^n}{\theta^n} & \text{si } 0 < s \leq \theta \\ 1 & \text{si } \theta < s \end{cases} \text{ respectivamente.}$$

Como $g_0(s, \theta)$ es creciente para cada s y $G_0(s, \theta)$ es continua en s y decreciente en θ , para encontrar el intervalo de probabilidad de S se tiene:

$$\begin{aligned} p(s_1(\theta) \leq S \leq s_2(\theta)) &= G_0(s_2(\theta)) - G_0(s_1(\theta)) \\ &= \frac{s_2^n(\theta)}{\theta^n} - \frac{s_1^n(\theta)}{\theta^n} \end{aligned}$$

$$\begin{aligned} p(s_1(\theta) \leq S \leq s_2(\theta)) &= \frac{1}{\theta^n} (s_2^n(\theta) - s_1^n(\theta)) \\ &= \gamma \end{aligned}$$

Como el límite superior de $g_0(s, \theta)$ es θ , se tiene $\frac{1}{\theta^n} (\theta^n - s_1^n(\theta)) = \gamma$, despejando de esta última ecuación a $s_1(\theta)$ queda:

$$s_1(\theta) = (1 - \gamma)^{1/n} \theta = \alpha^{1/n} \theta$$

Sustituyendo esta expresión en el intervalo de S , $p(\alpha^{1/n}\theta \leq S \leq \theta) = 1 - \alpha$ y $A(s_0) = \{\theta | \alpha^{1/n}\theta \leq s_0 \leq \theta\}$, por lo que el "intervalo fiducial" queda $\left(s_0, \frac{s_0}{\alpha^{1/n}}\right)$ con probabilidad fiducial $\gamma = 1 - \alpha$. El intervalo coincide con los resultados mostrados en el ejercicio 2.3 y su correspondiente en el capítulo 1.

Ejercicio 2.5.

Regresando a los datos del ejercicio 1.2. donde X_1, \dots, X_n son v.a. iid $f(x; \theta)$ con

$f(x; \theta) = \frac{3x^2}{\theta^3} I_{[0, \theta]}(x)$ y estadística suficiente minimal $S(\underline{X}) = \max\{X_1, \dots, X_n\}$ cuyas

funciones de densidad es $g_0(s; \theta) = \frac{3ns^{3n-1}}{\theta^{3n}} I_{[0, \theta]}(s)$ y función de distribución

acumulada $G_0(s, \theta) = \frac{s^{3n}}{\theta^{3n}} I_{[0, \theta]}(s) + I_{(\theta, \infty)}(s)$; sabiendo, además, que $g_0(s; \theta)$ es

creciente para s y $G_0(s; \theta)$ es continua en s y decreciente para θ .

El intervalo para S genera:

$$\begin{aligned} p(s_1(\theta) \leq S \leq s_2(\theta)) &= G_0(s_2(\theta)) - G_0(s_1(\theta)) \\ &= \frac{s_2^{3n}(\theta)}{\theta^{3n}} - \frac{s_1^{3n}(\theta)}{\theta^{3n}} \end{aligned}$$

$$\begin{aligned} p(s_1(\theta) \leq S \leq s_2(\theta)) &= \frac{1}{\theta^{3n}} (s_2^{3n}(\theta) - s_1^{3n}(\theta)) \\ &= \gamma \end{aligned}$$

El límite superior de $g_0(s; \theta)$ es θ por lo que,

$$\frac{1}{\theta^{3n}}(\theta^{3n} - s_1^{3n}(\theta)) = \gamma$$

$$s_1(\theta) = (1 - \gamma)\theta^{3n}$$

$$s_1(\theta) = \alpha\theta^{3n}$$

y

$p(\theta\alpha^{1/3n} \leq S \leq \theta) = 1 - \alpha$ y $A(s_0) = \{\theta | \alpha^{1/3n} \leq s_0 \leq \theta\}$ y el "intervalo fiducial" para θ quedaría $\left(s_0, \frac{s_0}{\alpha^{1/3n}}\right)$ con probabilidad fiducial $1 - \alpha$.

Ejemplo 2.6.

De los datos del ejercicio 1.7. donde X_1, \dots, X_n son v.a. iid $f(x; \theta)$, donde

$f(x; \theta) = \frac{xe^{-\frac{x^2}{2\theta}}}{\theta} I_{(0, \infty)}(x)$ y con $S(\underline{X}) = \sum_{i=1}^n X_i^2$ la estadística suficiente minimal la cual se

distribuye gama $\left(\lambda = \frac{1}{2\theta}, r = n\right)$ y con función de distribución

$$G_0(s, \theta) = \begin{cases} 0 & \text{si } s \leq 0 \\ \int_0^{\frac{s}{2\theta}} \frac{w^{n-1} e^{-w}}{\Gamma(n)} dw & \text{si } s > 0 \end{cases}$$

Como $f(s; \theta)$ es creciente para s y $G_0(s; \theta)$ es continua en s y decreciente para θ . El intervalo para S genera:

$$p(s_1(\theta) \leq S \leq s_2(\theta)) = G_0(s_2(\theta)) - G_0(s_1(\theta))$$

$$= \int_0^{\frac{s_2(\theta)}{2\theta}} \frac{w^{n-1} e^{-w}}{\Gamma(n)} dw - \int_0^{\frac{s_1(\theta)}{2\theta}} \frac{w^{n-1} e^{-w}}{\Gamma(n)} dw$$

$$= \gamma$$

El límite superior de $f(s; \theta)$ es ∞ por lo que,

$$\int_0^\infty \frac{w^{n-1} e^{-w}}{\Gamma(n)} dw - \int_0^{\frac{s_1(\theta)}{2\theta}} \frac{w^{n-1} e^{-w}}{\Gamma(n)} dw = \gamma$$

$$\frac{\Gamma(n)}{\Gamma(n)} - \int_0^{\frac{s_1(\theta)}{2\theta}} \frac{w^{n-1} e^{-w}}{\Gamma(n)} dw = \gamma$$

$$\int_0^{\frac{s_1(\theta)}{2\theta}} \frac{w^{n-1} e^{-w}}{\Gamma(n)} dw = 1 - \gamma = \alpha$$

Considerando $n = 4$ se tiene $e^{\frac{s_1(\theta)}{2\theta}} \left(-\frac{s_1^3(\theta)}{48\theta^3} - \frac{s_1^2(\theta)}{8\theta^2} - \frac{s_1(\theta)}{2\theta} - 6 \right) + 1 = 1 - \alpha$ que es igual al resultado obtenido en el ejemplo 1.7.

El siguiente ejemplo que se presenta y que proviene del capítulo anterior difiere de los anteriores porque no es un modelo de grupo; el intervalo fiducial será muy diferente al obtenido por una forma asintótica en ese capítulo 1, pero se verá con el procedimiento Bayesiano, que las funciones fiducial y posteriori son muy parecidas y por lo tanto el intervalo fiducial y el de máxima densidad posterior también son casi iguales.

Ejemplo 2.7.

Considerando los datos del ejercicio 1.9. donde X_1, \dots, X_n v. a. iid *Bernoulli*(p) y

considerando a la estadística suficiente minimal $S = \sum_{i=1}^n X_i \sim \text{Bin}(n, p)$, siendo

$G_0(s, p) = \sum \binom{n}{j} p^j (1-p)^{n-j}$, la cual es decreciente en p si $s = 0, 1, \dots, n-1$, por lo

$$\text{que } \frac{dG_0(s, p)}{dp} = -\frac{n!}{(n-s-1)!s!} p^s (1-p)^{n-s-1} \leq 0 \quad \forall p \in (0,1).$$

Como G_0 es monótona decreciente en p , excepto $s=n$, en este caso $G_0(n, p) = 1 \quad \forall p \in (0,1)$.

O'Reilly (2003) propone la distribución fiducial de p de modo que sea consistente respecto al concepto de "éxito" o "fracaso" (p o $1-p$), también puede verse como de s o su equivalente $n-s$, es decir,

$$F_{fid}(p) = \begin{cases} 1 - G_0(s, p) & \text{si } s = 0 \\ \frac{1}{2}(1 - G_0(s, p)) + \frac{1}{2}(1 - G_0^-(s, p)) & \text{si } s = 1, \dots, n-1 \\ 1 - G_0^-(s, p) & \text{si } s = n \end{cases}$$

Donde G_0^- indica que la función es continua por la izquierda.

La función de densidad queda:

$$f_{fid}(p) = \begin{cases} \text{Beta}(s+1, n-s) & \text{si } s = 0 \\ \frac{1}{2} \text{Beta}(s+1, n-s) + \frac{1}{2} \text{Beta}(s, n-s+1) & \text{si } s = 1, \dots, n-1 \\ \text{Beta}(s, n-s+1) & \text{si } s = n \end{cases}$$

Para obtener el intervalo fiducial central para θ , con "probabilidad fiducial" γ y basado en el valor observado $S = s_0$ (como parámetro), hay que encontrar cantidades p_1^* y p_2^* tales que $p_{fid}(p_1^* \leq p \leq p_2^*) = F_{fid}(p_2^*) - F_{fid}(p_1^*) = \gamma$, esto es

$$\begin{aligned} \frac{1}{2}(1-G_0(s_0, p_2^*)) + \frac{1}{2}(1-G_0^-(s_0, p_2^*)) - \left(\frac{1}{2}(1-G_0(s_0, p_1^*)) + \frac{1}{2}(1-G_0^-(s_0, p_1^*)) \right) &= \gamma \\ -\frac{1}{2}G_0(s_0, p_2^*) - \frac{1}{2}G_0^-(s_0, p_2^*) + \frac{1}{2}G_0(s_0, p_1^*) + \frac{1}{2}G_0^-(s_0, p_1^*) &= \gamma \\ G_0(s_0, p_1^*) - G_0(s_0, p_2^*) + G_0^-(s_0, p_1^*) - G_0^-(s_0, p_2^*) &= 2\gamma \end{aligned}$$

Existe un problema cuando $g(\theta)$ no es función monótona de θ . La función de distribución fiducial de $g(\theta)$ puede no ser la que resultaría al transformar la fiducial de θ , como se mostrará en el capítulo 3. Esto es un punto controversial que pudiera estar relacionado con una inferencia incorrecta.

2.2. Enfoque Bayesiano

Se considera a θ como variable aleatoria que tiene una distribución de probabilidad. La distribución de X depende de θ .

Se denota la función de probabilidad de θ como $\pi(\theta)$ para $\theta \in \Omega$.

Sea X_1, \dots, X_n una muestra aleatoria de la distribución de X y sea y una estadística función de las x 's.

Encontrando la función de probabilidad de y para θ dada, es decir $h(y/\theta)$; así la función de densidad conjunta de y y θ bajo condiciones muy generales es:

$$\pi(y; \theta) = \pi(\theta)h(y/\theta) \tag{31}$$

Si θ es variable aleatoria continua, entonces la función de densidad de probabilidad marginal de y es:

$$\pi(y) = \int_{-\infty}^{\infty} \pi(\theta)h(y/\theta)d\theta \quad (32)$$

Si θ es variable aleatoria discreta, entonces la función de densidad de probabilidad marginal de y es:

$$\pi(y) = \sum_{\theta} \pi(\theta)h(y/\theta) \quad (33)$$

Obteniendo ahora la función de densidad de θ/y se tiene:

$$\pi(\theta/y) = \frac{\pi(y,\theta)}{\pi(y)} = \frac{\pi(\theta)h(y/\theta)}{\pi(y)}; \pi(y) > 0 \quad (34)$$

$$\text{O como suele escribirse } \pi(\theta/y) \propto h(y/\theta)\pi(\theta) \quad (35)$$

En donde

$\pi(\theta)$ es llamada la distribución inicial de θ o distribución *a priori*.

$\pi(\theta/y)$ es llamada la distribución posterior (*final*) de θ o distribución *a posteriori*.

En el contexto Bayesiano no existe el concepto de intervalo de $(1-\alpha)\times 100\%$ de confianza para θ . Aquí se maneja el concepto de intervalo de probabilidad posterior $1-\alpha$ para θ , que posee propiedades óptimas como:

1. La distribución posterior para cada θ dentro del intervalo es al menos tan grande como la de los valores afuera del intervalo,
2. Para una probabilidad dada α el intervalo es lo más pequeño posible.

Bajo condiciones muy generales 1) implica 2).

Cuando un intervalo de probabilidad posterior posee las propiedades anteriores es llamado *intervalo de máxima densidad posterior* (*hpdi = highest posterior density interval*).

Traduciendo este concepto se dice que un intervalo es de máxima densidad posterior si:

Dada $\pi(\theta/y)$ de θ , la distribución posterior, y $\int_{\theta_1}^{\theta_2} \pi(\theta/y) d\theta = 1 - \alpha$, considerando

además que para θ' dentro del intervalo y θ'' fuera del intervalo de probabilidad, se tiene que $\pi(\theta'/y) \geq \pi(\theta''/y)$.

El intervalo se interpreta como: θ está entre θ_1 y θ_2 con una probabilidad final $1 - \alpha$; en el sentido formal probabilístico de Kolmogorov.

El principal detalle que tiene el procedimiento Bayesiano para el cálculo de intervalos de probabilidad es el supuesto del conocimiento inicial de θ , considerándose que esto puede obtenerse por experiencia previa o por argumentos teóricos; pero en algunas aplicaciones, no hay conocimiento inicial de θ . Existen algunos procedimientos que proponen la distribución *a priori* cuando no hay una apreciación subjetiva; entre otros se tienen:

- a) Distribuciones *a priori* conjugadas que surgen cuando $\pi(\theta/y)$ tendría una expresión matemática complicada; por lo que se intenta determinar una familia conjugada de distribuciones *a priori*, evitando esa dificultad y tal que: i) sea matemáticamente tratable, ii) incluya lo más posible distribuciones con distintas localizaciones, varianzas y formas para que puedan tratar diferentes estados iniciales de información y iii) sean fáciles de interpretar.

- b) Distribuciones *a priori* difusas, no informativas o de referencia. Esto se usa cuando se quiere seleccionar una $\pi(\theta)$ que proporcione poca información relativa a la información que se espera de la muestra X_1, \dots, X_n sobre el parámetro θ .
- c) Distribuciones *a priori* que reconozcan la verosimilitud trasladada por los datos. Es cuando se busca una transformación uno a uno $\phi(\theta)$ del parámetro θ , cuya verosimilitud sea aproximadamente conocida *a priori* excepto por su localización, la cual será determinada por la muestra X_1, \dots, X_n . Esta consideración conduce a la famosa distribución *a priori* propuesta por Jeffrey's que en el caso escalar es $\pi_J(\theta) \propto |i_E(\theta)|^{1/2}$ (Box y Tiao 1992), con $i_E(\theta)$ la información esperada de Fisher.

Es importante señalar, que a veces al aplicar estos procedimientos para obtener la distribución inicial $\pi(\theta)$, ésta resulta ser impropia, es decir,

$$\begin{aligned} \sum \pi(\theta) &= \infty & \text{si } \theta \text{ es discreta} \\ \int \pi(\theta) d\theta &= \infty & \text{si } \theta \text{ es continua} \end{aligned} \tag{36}$$

Pero si la distribución posterior resulta propia suele no haber mayores problemas en utilizar distribuciones *a priori* impropias, pero no siempre es el caso. Este tema no se trata en este trabajo.

En relación a los parecidos que pudiera haber entre los resultados bayesianos y los obtenidos por la estadística clásica, Lindley (1958), Welch y Peers (1963) mostraron que los intervalos de máxima densidad posterior $1-\alpha$ obtenidos a partir de distribuciones *a priori* no informativas, en muchas ocasiones coinciden numéricamente con los intervalos

de confianza más pequeños para θ al nivel $1-\alpha$.

Ejemplo 2.8.

Considerando la información del ejercicio 2.1. donde X_1, \dots, X_n son v. a. iid $Exp(\lambda)$, supóngase una distribución *a priori* no informativa impropia, $\pi(\lambda) = \frac{1}{\lambda}$; $\lambda \in (0, \infty)$, entonces como la verosimilitud de λ es proporcional a la función de densidad de la estadística suficiente minimal $\left(s(\underline{X}) = \sum_{i=1}^n X_i \right)$, entonces la distribución posterior de λ se puede calcular como:

$$\pi(\lambda / y) = \frac{\pi(\lambda)h(y/\lambda)}{\pi(y)}, \text{ donde } h(y/\lambda) = g_0(s; \lambda) = \begin{cases} \frac{\lambda^n s^{n-1} e^{-\lambda s}}{\Gamma(n)} & \text{si } s \in (0, \infty) \\ 0 & \text{eoc} \end{cases}$$

$$\pi(\lambda)h(y/\lambda) = \begin{cases} \frac{\lambda^{n-1} s^{n-1} e^{-\lambda s}}{\Gamma(n)} & \text{si } \lambda \in (0, \infty) \\ 0 & \text{eoc} \end{cases}$$

$$\begin{aligned} \pi(y) &= \int_0^\infty \frac{\lambda^{n-1} s^{n-1} e^{-\lambda s}}{\Gamma(n)} d\lambda \\ &= \frac{s^{-1}}{\Gamma(n)} \int_0^\infty w^{n-1} e^{-w} dw \\ &= \frac{1}{s} \end{aligned}$$

$$\pi(\lambda / y) = \pi(\lambda / s) = \begin{cases} \frac{\lambda^{n-1} s^n e^{-\lambda s}}{\Gamma(n)} & \text{si } 0 < \lambda < \infty \\ 0 & \text{e.o.c.} \end{cases}$$

Para encontrar el intervalo de máxima densidad posterior de contenido $1-\alpha$ se debe encontrar las constantes $c_{1\alpha}$ y $c_{2\alpha}$, tal que

$$\int_{c_{1\alpha}}^{c_{2\alpha}} \pi(\lambda / y) d\lambda = 1 - \alpha$$

$$\int_{c_{1\alpha}}^{c_{2\alpha}} \frac{\lambda^{n-1} s^n e^{-\lambda s}}{\Gamma(n)} d\lambda = 1 - \alpha$$

Esta integral queda: $\frac{1}{\Gamma(n)} \int_{c_{1\alpha}}^{c_{2\alpha}} w^{n-1} e^{-w} dw = 1 - \alpha$, quedando igual que los ejemplos 2.1. y

1.3.

Ejemplo 2.9.

Retomando el ejercicio 2.3., donde se tenía una muestra X_1, \dots, X_n v.a. iid $U(0, \theta)$, se supone una distribución *a priori* propia para θ como $\pi(\theta) = 2\theta$; $\theta \in (0, 1)$, como la verosimilitud de θ es proporcional a la función de densidad de la estadística suficiente minimal ($S = \text{máx}\{X_1, \dots, X_n\}$), entonces la distribución posterior de θ se puede calcular como:

$$\pi(\theta / y) = \frac{\pi(\theta)h(y/\theta)}{\pi(y)}, \text{ donde } h(y/\theta) = g_0(s; \theta) = \begin{cases} \frac{ns^{n-1}}{\theta^n} & \text{si } s \in [0, 1] \\ 0 & \text{eoc} \end{cases}$$

$$\pi(\theta)h(y/\theta) = \begin{cases} \frac{2ns^{n-1}}{\theta^{n-1}} & \text{si } \theta \in [s, 1] \\ 0 & \text{eoc} \end{cases}$$

$$\pi(y) = \int_s^1 \frac{2ns^{n-1}}{\theta^{n-1}} d\theta = \frac{2ns^{n-1}}{2-n} \theta^{-n+2} \Big|_s^1 = \frac{2ns}{n-2} (1 - s^{n-2})$$

$$\pi(\theta/y) = \pi(\theta/s) = \begin{cases} \frac{(n-2)s^{n-2}}{(1-s^{n-2})\theta^{n-1}} & \text{si } s \leq \theta \leq 1 \\ 0 & \text{e.o.c.} \end{cases}$$

Para encontrar el intervalo de máxima densidad posterior de contenido $1-\alpha$ se debe encontrar la constante c_α , tal que

$$\int_s^{c_\alpha} \pi(\theta/y) d\theta = 1-\alpha$$

$$\int_s^{c_\alpha} \frac{(n-2)s^{n-2}}{(1-s^{n-2})\theta^{n-1}} d\theta = 1-\alpha$$

Al resolver la integral deja a c_α como:

$$c_\alpha = s \left(s^{n-2} (1-\alpha) + \alpha \right)^{\frac{1}{n-2}}$$

Así el intervalo de máxima densidad posterior para θ queda:

$$\left(s, s \left(s^{n-2} (1-\alpha) + \alpha \right)^{\frac{1}{n-2}} \right)$$

Este intervalo difiere al encontrado en el ejemplo 2.3., pero si se cambia la distribución *a priori* por una no informativa impropia éste cambiará.

Supóngase ahora que $\pi(\theta) = \frac{1}{\theta}$; $\theta \in (0, \infty)$, entonces:

$$\pi(\theta)h(y/\theta) = \begin{cases} \frac{ns^{n-1}}{\theta^{n+1}} & \text{si } \theta \in (s, \theta) \text{ y} \\ 0 & \text{e.o.c.} \end{cases}$$

$$\begin{aligned}\pi(y) &= \int_s^\infty \frac{ns^{n-1}}{\theta^{n+1}} d\theta \\ &= \frac{ns^{n-1}}{-n} \theta^{-n} \Big|_s^\infty \\ \pi(y) &= \frac{1}{s}\end{aligned}$$

Por lo que la distribución posterior de θ es igual a

$$\pi(\theta/y) = \pi(\theta/s) = \begin{cases} \frac{ns^n}{\theta^{n+1}} & \text{si } \theta \in (s, \infty) \\ 0 & \text{e.o.c.} \end{cases}$$

Haciendo lo mismo para obtener la constante c_α

$$\begin{aligned}\int_s^{c_\alpha} \pi(\theta/y) d\theta &= 1 - \alpha \\ \int_s^{c_\alpha} \frac{ns^n}{\theta^{n+1}} d\theta &= 1 - \alpha\end{aligned}$$

Al resolver la integral queda: $c_\alpha = \frac{s}{\alpha^{1/n}}$, así el intervalo de máxima densidad posterior

para θ es: $\left(s, \frac{s}{\alpha^{1/n}}\right)$, que es el mismo que se obtuvo en el ejemplo 2.3. y el correspondiente al capítulo uno (ejemplo 1.6.).

Ejemplo 2.10.

Sea X_1, \dots, X_n una muestra aleatoria de *Bernoulli*(p); $0 < p < 1$

Sea $Y = \sum_{i=1}^n X_i$ la estadística suficiente minimal de p

$$h(y/p) = \text{Bin}(n, p)$$

$$\Sigma y_i \sim \text{Bin}(n, \theta)$$

$$h(y/p) = \begin{cases} \binom{n}{y} p^y (1-p)^{n-y} & \text{si } y = 0, 1, \dots, n \\ 0 & \text{si } \text{eoc} \end{cases}$$

Tomando ahora la distribución Beta para p esto es:

$$\pi(p) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1} (1-p)^{b-1}; 0 < p < 1, a, b \text{ constantes positivas.}$$

$$\begin{aligned} \pi(y) &= \int_0^1 \frac{\Gamma(a+b) p^{a-1} (1-p)^{b-1} \binom{n}{y} p^y (1-p)^{n-y}}{\Gamma(a)\Gamma(b)} dp \\ &= \frac{\binom{n}{y} \Gamma(a+b) \Gamma(a+y) \Gamma(b+n-y)}{\Gamma(a)\Gamma(b)\Gamma(a+b+n)} \quad \text{si } y = 0, 1, \dots, n \end{aligned}$$

Obteniendo ahora la distribución posterior de p/y .

$$\pi(p/y) = \frac{\Gamma(a+b) p^{a-1} (1-p)^{b-1} \binom{n}{y} p^y (1-p)^{n-y}}{\Gamma(a)\Gamma(b)}; \quad 0 < p < 1$$

$$\frac{\binom{n}{y} \Gamma(a+b) \Gamma(a+y) \Gamma(b+n-y)}{\Gamma(a)\Gamma(b)\Gamma(a+b+n)}$$

$$\pi(p/y) = \begin{cases} \frac{p^{a+y-1} (1-p)^{b+n-y-1}}{B(a+y; b+n-y)} & \text{si } 0 < p < 1 \\ 0 & \text{si } eoc \end{cases}$$

$$= \text{Beta}(A = a+y, C = b+n-y)$$

Para encontrar el intervalo de máxima densidad posterior de contenido $1-\alpha$ se debe encontrar las constantes $c_{1\alpha}$ y $c_{2\alpha}$, tal que

$$\int_{c_{1\alpha}}^{c_{2\alpha}} \pi(p/y) dp = 1-\alpha$$

$$\int_{c_{1\alpha}}^{c_{2\alpha}} \frac{p^{a+y-1} (1-p)^{b+n-y-1}}{B(a+y; b+n-y)} dp = 1-\alpha$$

Como puede observarse no coincide la solución de esta integral con el intervalo construido en el ejercicio 1.9, por lo que ahora se procederá a mostrar que sí pueden casi coincidir bajo ciertas consideraciones.

Tomando, un caso particular, como inicial ¹⁰

$$\pi(p) = p^{-\frac{1}{2}} (1-p)^{-\frac{1}{2}}; \quad 0 < p < 1,$$

se llega a que:

¹⁰ Se le conoce como distribución a priori de Jeffrey's.

$$\begin{aligned}\pi(y) &= \int_0^1 p^{-\frac{1}{2}} (1-p)^{-\frac{1}{2}} \binom{n}{y} p^y (1-p)^{n-y} dp \\ &= \frac{\binom{n}{y} \Gamma\left(y + \frac{1}{2}\right) \Gamma\left(n - y + \frac{1}{2}\right)}{\Gamma(n+1)} \quad \text{si } y = 0, 1, \dots, n\end{aligned}$$

y por tanto

$$\begin{aligned}\pi(p/y) &= \begin{cases} \frac{p^{y-\frac{1}{2}} (1-p)^{n-y-\frac{1}{2}}}{B\left(y + \frac{1}{2}; n - y + \frac{1}{2}\right)} & \text{si } 0 < p < 1 \\ 0 & \text{si } eoc \end{cases} \\ &= \text{Beta}\left(A = y + \frac{1}{2}, C = n - y + \frac{1}{2}\right)\end{aligned}$$

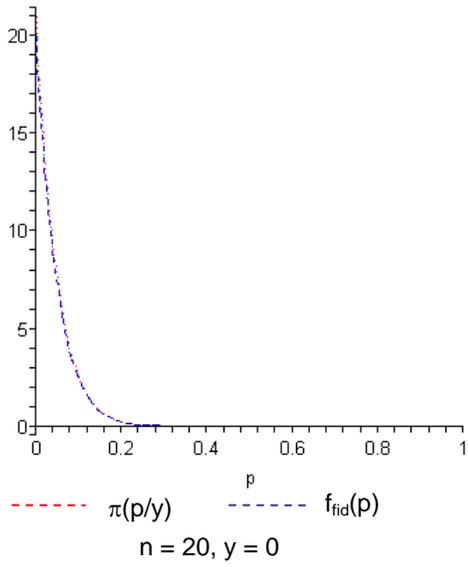
Puede notarse que se parece mucho a la densidad fiducial dada en el ejemplo 2.7.

Para encontrar el intervalo de máxima densidad posterior de contenido $1-\alpha$ se debe encontrar las constantes $c_{1\alpha}$ y $c_{2\alpha}$, tal que

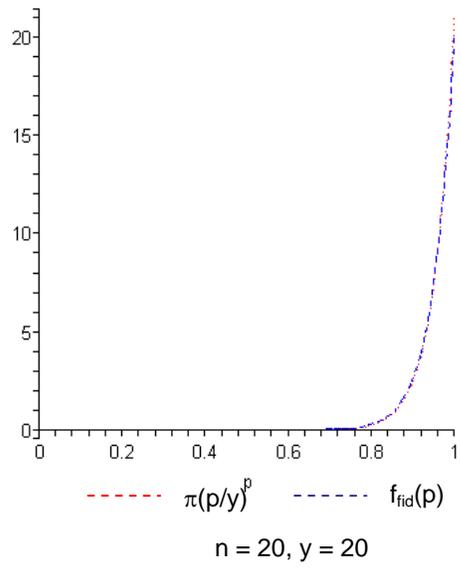
$$\begin{aligned}\int_{c_{1\alpha}}^{c_{2\alpha}} \pi(p/y) dp &= 1 - \alpha \\ \int_{c_{1\alpha}}^{c_{2\alpha}} \frac{p^{y-\frac{1}{2}} (1-p)^{n-y-\frac{1}{2}}}{B\left(y + \frac{1}{2}; n - y + \frac{1}{2}\right)} dp &= 1 - \alpha\end{aligned}$$

A continuación se muestran las gráficas de las funciones fiducial y posterior para $n = 20$ y como podrá apreciarse en ellas el parecido es muy grande, tal que materialmente son la misma función.

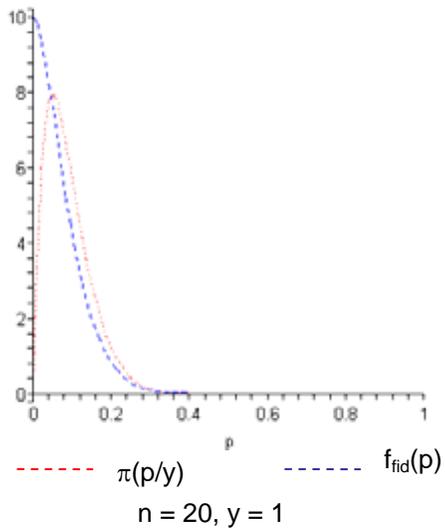
Gráfica 2.1.



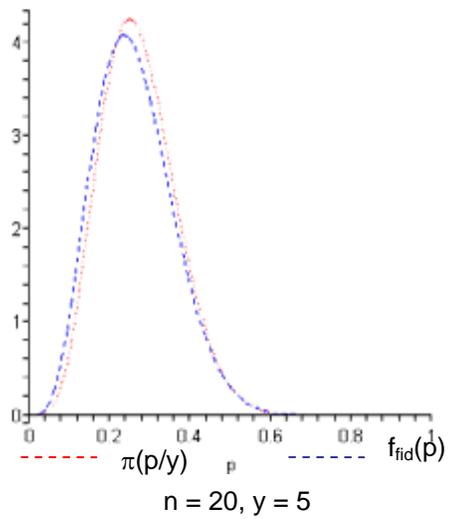
Gráfica 2.2.



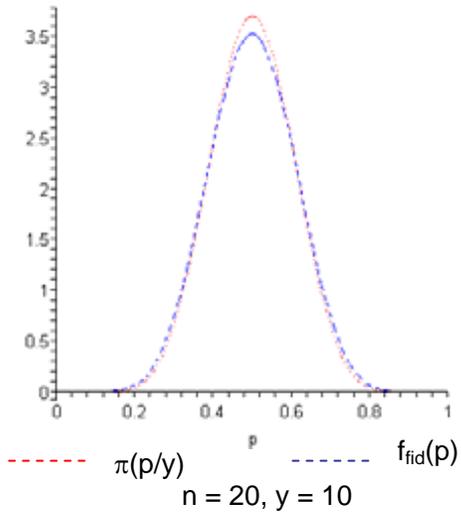
Gráfica 2.3.



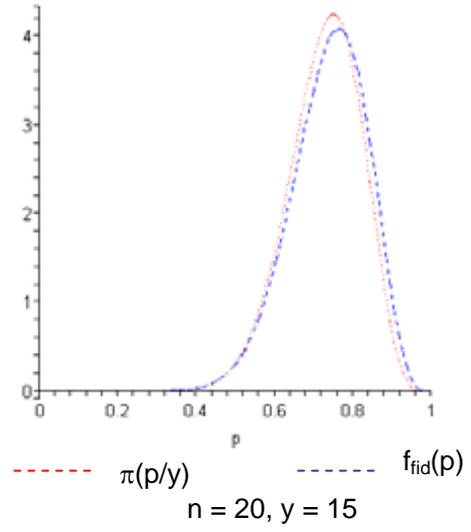
Gráfica 2.4.



Gráfica 2.5.



Gráfica 2.6.



Capítulo 3

Distribuciones fiducial y posterior. Semejanzas y diferencias

Introducción

En este capítulo se mostrarán las características básicas de la distribución fiducial, que de alguna manera fueron revisadas en el capítulo anterior; en esta parte se señalada el resultado de que, para el caso de un parámetro real, cualquier familia, miembro de la familia exponencial cumple los criterios de contrastabilidad. Posteriormente se examinará con ejemplos, cómo la función de distribución fiducial puede aproximarse mucho a la distribución posterior en su forma numérica. A las distribuciones fiduciales que generan una masa en el extremo se les ha llamado defectuosas, en la literatura, (en el sentido de que la parte de la densidad que se comporta continua, no integra a uno), pero esto no le impide que su aproximación a la distribución posterior, cuando no hay masa, calculada ésta con la llamada inicial no-informativa, sea muy buena, de tal forma que no hay en la práctica gran diferencia.

3.1. Características básicas de la distribución fiducial

De lo visto en el capítulo anterior se puede resumir lo siguiente:

La verosimilitud para θ es proporcional a $g_0(s; \theta)$, la densidad de S , toda la información necesaria para inferir sobre θ está contenida en $G_0(s; \theta)$, función de distri-

bución de la estadística suficiente minimal.

Hay una relación entre S y θ en el sentido de que valores grandes de θ implican valores grandes de S y los valores pequeños de θ implican valores "estadísticamente pequeños" para S .

Esto se usa, por ejemplo cuando para

$$H_0 : \theta \leq \theta_0 \text{ vs } H_1 : \theta > \theta_0 \quad (1)$$

"se rechaza H_0 para valores grandes de S " (puesto que valores pequeños del parámetro son los asociados a valores pequeños de la estadística).

Al fijar la hipótesis $H_0 : \theta \leq \theta_0$, entonces

$$1 - G_0(s; \theta_0) = P(S > s; \theta_0) \quad (2)$$

será una cantidad decreciente en s , ya que manteniendo la hipótesis fija, la evidencia al hacerse más extrema en contra de H_0 produce un nivel de significancia descriptivo (valor p), más pequeño.

Análogamente se podría suponer que s es fija y θ_0 se hace variar. El nivel de significancia descriptivo para $H_0 : \theta \leq \theta_0$ es

$$1 - G_0(s; \theta_0), \quad (3)$$

y “debiera” ocurrir que al tomar $\theta'_0 < \theta_0$, el correspondiente $1 - G_0(s; \theta'_0)$ debería ser más chico. Esto implicaría que $1 - G_0(s; \theta)$ debe ser monótona decreciente como función de θ , para cada s fija.

A esta propiedad se le llama contrastabilidad monótona entre la evidencia y la hipótesis (O'Reilly, 2003) y corresponde a otras nociones mencionadas en la literatura.

En conclusión, es claro que cuando θ_0 es fija en $H_0: \theta \leq \theta_0$, se considera que al converger s a su límite superior, el valor de significancia descriptivo se aproxima a cero; pero cuando s es la fija, no es evidente que cuando θ_0 converja a su límite inferior, el correspondiente valor de significancia descriptivo tienda a cero.

Esto se puede resumir en la siguiente definición.

Definición 3.1. Contrastabilidad Monótona.

Hay ***contrastabilidad monótona*** si $G_0(s; \theta)$ es decreciente como función de θ ;

Existe ***certidumbre en la contrastabilidad*** para casos extremos si se cumple:

$$G_0(s; \theta) \rightarrow 1 \text{ si } \theta \rightarrow \text{límite inferior,}$$

$$G_0(s; \theta) \rightarrow 0 \text{ si } \theta \rightarrow \text{límite superior.}$$

3.2. Distribución fiducial

Si $G_0(s; \theta)$ satisface la contrastabilidad monótona, entonces la función de distribución fiducial es simplemente:

$$F_{fid}(\theta; s) = 1 - G_0(s; \theta) \quad (4)$$

por esto F_{fid} será creciente en θ .

Si la contrastabilidad es extrema o la certidumbre en la contrastabilidad es satisfecha, entonces la distribución fiducial no tiene una "masa fiducial"¹¹ en los puntos finales del intervalo θ .

El uso de la distribución fiducial es directo y provee intervalos de confianza (fiduciales) igual si F_{fid} tiene una "singularidad" o "masa fiducial" en un punto extremo de θ (O'Reilly, 2003).

3.3. Revisión del cumplimiento de la contrastabilidad monótona

Si la función de distribución de la estadística suficiente cumple el criterio de contrastabilidad monótona, se tiene que los intervalos de confianza y los intervalos fiduciales producen los mismos resultados, como ejemplo de esto se da el siguiente teorema (O'Reilly, 2003) que abarca a la familia exponencial natural ¹².

¹¹ La masa fiducial es una acumulación de "densidad" en un punto, más adelante se hace referencia a este concepto.

¹² $g_0(s; \theta)$ es una función de densidad que pertenece a la familia exponencial natural si $g_0(s; \theta) = c(s) e^{\theta s - M(\theta)}$, para $s \in R_s$, donde R_s es un intervalo abierto que no depende de θ y $\theta \in \Theta$, donde

Θ es un intervalo y $M(\theta)$ es el cumulante $M(\theta) = \log \left\{ \int_{R_s} c(s) e^{\theta s} ds \right\}$.

Teorema 3.1¹³

Si $g_0(s; \theta)$ es un miembro de la familia exponencial natural; entonces $G_0(s; \theta)$ es monótona decreciente en θ , para toda $s \in \mathcal{R}_s$.

Con este resultado un miembro de la familia exponencial natural permite siempre inducir una distribución fiducial. Ésta no tendrá una “masa fiducial” en los extremos del intervalo paramétrico si la familia es regular (en el sentido de que el conjunto de valores del parámetro para los cuales la integral de la densidad es finita, es abierto; o sea que en los valores extremos la distribución es degenerada).

Ejemplo 3.1

Considerando los datos dados en el ejemplo 2.1. donde X_1, \dots, X_n constituye una muestra aleatoria de una $Exp(\lambda)$, y la estadística suficiente minimal es $S(\underline{X}) = \sum_{i=1}^n X_i$, la cual se distribuye gama $\left(\alpha = n, \beta = \frac{1}{\lambda}\right)$; por lo anterior expuesto $g_0(s; \lambda)$ cumple con la contrastabilidad monótona y la contrastabilidad extrema.

También se puede expresar a la función de distribución de S , como:

$$F(s; \lambda, n) = \int_0^s \frac{\lambda^n u^{n-1} e^{-\lambda u}}{\lambda(n)} du \quad (5)$$

¹³ La demostración del teorema puede verse en O'Reilly.

Con un cambio de variable $w = \lambda u$, se obtiene

$$F(s; n) = \int_0^{su} \frac{w^{n-1} e^{-w}}{\lambda(n)} dw \quad (6)$$

Forma que coincide con los ejemplos mostrados en 1.1 y 2.1 y por lo tanto cumple con lo antes referido. Lo anterior fue expresado por Lindley (1961) con los siguientes resultados.

Resultado 3.1.

Una condición necesaria y suficiente para que la distribución fiducial de θ dada x sea la distribución de probabilidad final (Bayes), es que exista una transformación de x a u y de θ a τ , tal que τ es un parámetro de localización de u .

Resultado 3.2.

Si para una muestra aleatoria de cualquier tamaño de la distribución de x , tal que existe una sola estadística suficiente de θ , entonces el argumento fiducial es inconsistente, a menos que el Resultado 3.1 se obtenga, y si es así, el argumento fiducial es equivalente al argumento Bayesiano con una distribución *a-priori* uniforme para τ .

3.4. Ejemplos donde el parámetro no es de localización o hay masa en alguno de los extremos del intervalo del parámetro.

Con los resultados 3.1. y 3.2. se indican las condiciones en las que las distribuciones fiducial y final coincidirán. Lo interesante es ahora revisar qué sucede cuando las

condiciones que propone Lindley no se cumplen; ya se sabe, entonces que no habrá esa coincidencia en las funciones pero numéricamente serán muy parecidas.

Ejemplo 3.2.

Considerando a una función de densidad de probabilidad gama:

$$f(x; \theta) = \frac{x^{\theta-1} e^{-x}}{\Gamma(\theta)}; \quad x > 0, \theta > 0 \quad (7)$$

Con función de distribución

$$F(x; \theta) = \int_0^x \frac{u^{\theta-1} e^{-u}}{\Gamma(\theta)} du = \frac{1}{\Gamma(\theta)} \int_0^x u^{\theta-1} e^{-u} du \quad (8)$$

Al integrar por partes esta función puede ser expresada como:

$$F(x; \theta) = -\frac{x^{\theta-1} e^{-x}}{\Gamma(\theta)} + \frac{1}{\Gamma(\theta-1)} \int_0^x u^{\theta-2} e^{-u} du \quad (9)$$

$$F(x; \theta) = -f(x; \theta) + F(x; \theta-1)$$

Para revisar que cumple el criterio de contrastabilidad monótona, se propone

$\theta' = \theta - 1 > 0$, $\theta = \theta' + 1$ por lo que $\theta' < \theta$; considerando

$$F(x; \theta' + 1) = -f(x; \theta' + 1) + F(x; \theta' + 1 - 1)$$

$$= -f(x; \theta' + 1) + F(x; \theta')$$

\therefore

$$F(x; \theta') = F(x; \theta' + 1) + f(x; \theta' + 1) \quad (10)$$

\therefore

$$F(x; \theta') > F(x; \theta)$$

$\therefore F(x; \theta)$ es monótona decreciente para θ , para cualquier $x > 0$.

Para la contrastabilidad en casos extremos se requiere tomar los límites en los extremos del intervalo de variación de θ .

$$\begin{aligned}\lim_{\theta' \rightarrow 0} F(x; \theta') &= \lim_{\theta' \rightarrow 0} F(x; \theta' + 1) + \lim_{\theta' \rightarrow 0} F(x; \theta' + 1) \\ &= F(x; 1) + f(x; 1) \\ &= -e^{-x} + 1 + e^{-x} \\ &= 1\end{aligned}$$

Considerando (9) se puede seguir integrando por partes llegando a la fórmula:

$$F(x; \theta) = -\sum_{i=1}^{\theta-1} \frac{x^{\theta-i} e^{-x}}{\Gamma(\theta-i+1)} + \int_0^x e^{-u} du \quad (11)$$

por lo que tomando el límite de θ' hacia el extremo superior queda:

$$\begin{aligned}\lim_{\theta' \rightarrow \infty} F(x; \theta') &= \lim_{\theta' \rightarrow \infty} \left(-\sum_{i=1}^{\theta-1} \frac{x^{\theta-i} e^{-x}}{\Gamma(\theta-i+1)} \right) + \lim_{\theta' \rightarrow \infty} \int_0^x e^{-u} du \\ \lim_{\theta' \rightarrow \infty} F(x; \theta') &= -1 + e^{-x} - e^{-x} + 1 = 0\end{aligned}$$

Por lo que no se produce una "masa fiducial" en $\{0\}$ o cuando tiende a ∞ .

Obsérvese sin embargo que esta distribución no es un caso en que θ o una función de ella sea de localización; es decir, no cumple con las condiciones de Lindley, pero las funciones de probabilidad fiducial y de probabilidad final son muy parecidas como puede verse con lo siguiente:

Obteniendo la función de probabilidad fiducial para θ , habiendo observado x es:

$$\begin{aligned}
 f_{fid}(\theta; x) &= -\frac{d}{d\theta} F(x; \theta) \\
 &= -\frac{d}{d\theta} \left(\frac{1}{\Gamma(\theta)} \int_0^x u^{\theta-1} e^{-u} du \right) \\
 &= -\int_0^x e^{-u} \left(\frac{u^{\theta-1} \log u}{\Gamma(\theta)} - \frac{u^{\theta-1} \Gamma'(\theta)}{\Gamma^2(\theta)} \right) du \quad (12) \\
 &= \int_0^x \frac{u^{\theta-1} e^{-u}}{\Gamma(\theta)} (\psi(\theta) - \log u) du \\
 &= \int_0^x (\psi(\theta) - \log u) f(u; \theta) du
 \end{aligned}$$

Donde $\psi(\theta)$ es la función digama ¹⁴.

La estadística suficiente minimal para θ es $S = \prod_{i=1}^n X_i$ y por Lomnicki (1967) su función de densidad de probabilidad es:

$$g_0(s; \theta) = \frac{s^{\theta-1} \Phi_n(x)}{\Gamma^n(\theta)}, \text{ donde } \Phi_n(s) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} s^{-t} \Gamma^n(t) dt. \text{ Así la distribución posterior}$$

queda:

¹⁴ La función digama es la siguiente:

$$\psi(\theta) = \frac{d \log \Gamma(\theta)}{d\theta} = \frac{\Gamma'(\theta)}{\Gamma(\theta)}, \text{ donde } \Gamma(\theta) \text{ la función gama, } \Gamma(\theta) = \frac{e^{-\gamma\theta}}{\theta} \prod_{n=1}^{\infty} \left(1 + \frac{\theta}{n}\right)^{-1} e^{\frac{\theta}{n}}, \gamma \text{ es la constante}$$

de Euler-Mascheroni. Se puede desarrollar a la función digama y su expresión sería:

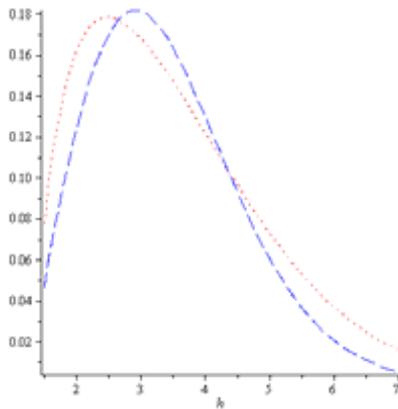
$$\psi(\theta) = -\gamma - \frac{1}{\theta} + \sum_{n=1}^{\infty} \left(\frac{1}{n} - \frac{1}{n+\theta} \right) = -\gamma + \sum_{k=1}^{\infty} \left(\frac{1}{k} - \frac{1}{\theta+k-1} \right).$$

Además $\gamma = \lim_{n \rightarrow \infty} \left[\sum_{k=1}^n \frac{1}{k} - \log n \right] = \int_1^{\infty} \left(\frac{1}{\lfloor x \rfloor} - \frac{1}{x} \right) dx$ aproximadamente tiene un valor de 0.577215664901532860606.

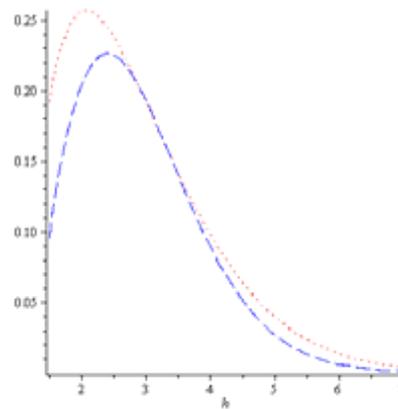
$$\pi(\theta / y) = \pi(\theta / s) = \frac{\sqrt{\psi(\theta)}s^{\theta-1}}{\Gamma^n(\theta) \int_0^\infty \frac{\sqrt{\psi(\theta)}s^{\theta-1}}{\Gamma^n(\theta)} d\theta} \quad (13)$$

Se hicieron las representaciones físicas para un par de evaluaciones de $S = 3$ y $S = 2$, las gráficas 3.1 y 3.2 indican con la forma punteada se muestra la distribución fiducial dada en (12) y la segmentada representa la distribución posterior dada en (13). Puede verse que las dos distribuciones son parecidas en términos gráficos, aunque no lo sean algebraicamente¹⁵.

Gráfica 3.1



Gráfica 3.2



Ejemplo 3.3. Distribución exponencial truncada.

En O'Reilly y Rueda (2006) se expone el ejemplo de la función exponencial truncada, donde se muestra que la distribución fiducial es muy parecida a la función de probabilidad final. A continuación se muestran estas semejanzas.

¹⁵ Las gráficas muestran en el eje de las x, la dependencia de θ , pero sólo por hacer la aplicación numérica se utilizó la letra h.

La distribución exponencial truncada se obtiene por condicionar la variable aleatoria exponencial X en que varíe solamente en el intervalo unitario, esto también hace que el parámetro original de la función pueda ser negativo en esta distribución condicional.

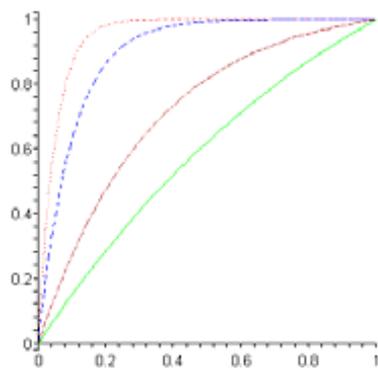
Así la familia de distribuciones es:

$$F(x; \theta) = \frac{e^{\theta x} - 1}{e^{\theta} - 1}, \quad x \in (0,1), \theta \in \mathbb{R} \quad (14)$$

Como puede verse en las gráficas 3.3., 3.4., 3.5. y 3.6. para valores de θ negativos las distribuciones tienen su masa cargada más hacia la primera mitad del intervalo unitario; para valores positivos de θ las distribuciones correspondientes tienen su masa hacia la segunda mitad del intervalo unitario; si se revisa en los extremos, cuando θ se aproxima a su límite superior (∞), la distribución se degenera a cero y si θ se aproxima a su límite inferior la degeneración es a uno. En el punto medio, cuando $\theta = 0$, le corresponde una distribución uniforme.

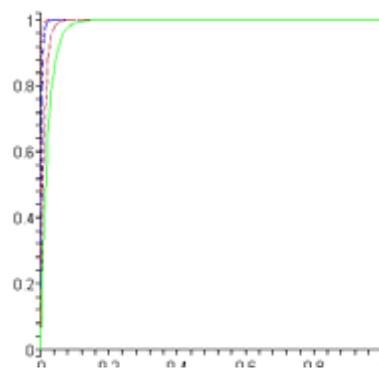
Gráfica 3.3

Función de distribución exponencial truncada, cuando $\theta < 0$



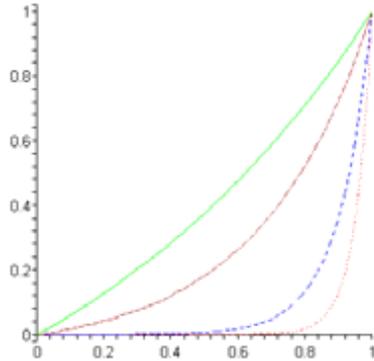
Gráfica 3.4

Función de distribución exponencial truncada, cuando $\theta < 0$



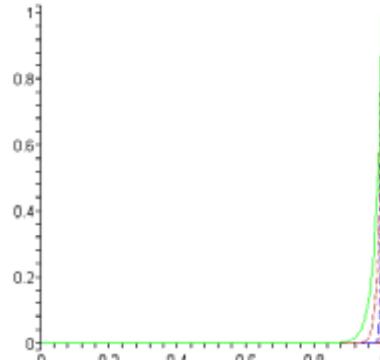
Gráfica 3.5

Función de distribución exponencial truncada, cuando $\theta > 0$



Gráfica 3.6

Función de distribución exponencial truncada, cuando $\theta > 0$



La estadística suficiente minimal es $S(\underline{X}) = \sum_{i=1}^n X_i$, en el caso de que $\theta = 0$, Feller (1985) da la densidad para esta suma cuando las X_1, \dots, X_n se distribuyen uniformes en $(0,1)$,

$$h_n(s) = \frac{1}{(n-1)!} \sum_{j=0}^{n-1} (-1)^j \binom{n}{j} (s-j)_+^{n-1}, \quad s \in (0, n) \quad 16$$

(15)

O'Reilly y Rueda calculan la función de densidad de S , para una θ arbitraria y la distribución fiducial, así

$$g_0(s; \theta) = \left(\frac{\theta}{e^\theta - 1} \right) e^{\theta s} h_n(s), \quad (16)$$

así que la distribución de S queda como:

¹⁶ $(x_+)^n$ es una función que es igual a 0 para $x \leq 0$ y es igual a x^n cuando $x \geq 0$, $(x-a)_+$ es cero para $x < a$ y $x-a$ cuando $x > a$.

Feller, W. (1985). (Teoría de las probabilidades y sus aplicaciones). Aclara que la suma no llega hasta n , ya que para $j = n$ y $s < n$ y n conocida, el último término es cero si $n > 1$.

$$G_0(s; \theta) = \int_0^s \left(\frac{\theta}{e^\theta - 1} \right)^n e^{\theta u} h_n(u) du \quad (17)$$

Los autores indican que por ser miembro de la familia exponencial natural, entonces se cumple el teorema 1 y por lo tanto, la distribución fiducial de θ será:

$$F_{fid}(\theta; s) = 1 - G_0(s; \theta)$$

$$F_{fid}(\theta; s) = 1 - \int_0^s \left(\frac{\theta}{e^\theta - 1} \right)^n e^{\theta u} h_n(u) du \quad (18)$$

Por lo que la densidad fiducial de θ será

$$f_{fid}(\theta; s) = \int_0^s \left(\frac{\theta}{e^\theta - 1} \right)^n \left(\frac{n(e^\theta(\theta - 1) - 1)}{\theta(e^\theta - 1)} - u \right) e^{\theta u} h_n(u) du$$

$$= \left(\frac{\theta^n}{e^\theta - 1} \right)^n \int_0^s \left(\frac{n(e^\theta(\theta - 1) - 1)}{\theta(e^\theta - 1)} - u \right) e^{\theta u} h_n(u) du \quad (19)$$

Esta densidad fiducial no tiene una expresión sencilla, pero puede ser evaluada numéricamente.

Calculando ahora la densidad posterior de probabilidad y considerando la distribución *a priori* de Jeffrey's como:

$$\pi(\theta) \propto \left(\frac{1}{\theta^2} + \frac{1}{2 - (e^\theta + e^{-\theta})} \right)^{1/2} \quad (20)$$

y con la estadística suficiente minimal $S(\underline{X}) = \sum_{i=1}^n X_i$, entonces la función de probabilidad *posterior* de θ se calcula como:

$$\pi(\theta|y) = \frac{\pi(\theta)h(y|\theta)}{\pi(y)}, \quad (21)$$

donde $h(y|\theta) = g_0(s;\theta)$, dada en (16):

$$\pi(\theta)h(y|\theta) = \left(\frac{1}{\theta^2} + \frac{1}{2-(e^\theta + e^{-\theta})} \right)^{1/2} \left(\frac{\theta}{e^\theta - 1} \right)^n e^{\theta s} h_n(s) \quad (22)$$

Considerando a $\pi(y) = \int_{-\infty}^{\infty} \pi(\theta)h(y|\theta)d\theta$, entonces queda:

$$\pi(y) = \int_{-\infty}^{\infty} \left(\frac{1}{\theta^2} + \frac{1}{2-(e^\theta + e^{-\theta})} \right)^{1/2} \left(\frac{\theta}{e^\theta - 1} \right)^n e^{\theta s} h_n(s) d\theta, \text{ excepto para } \theta = 0; \text{ así que la}$$

función de probabilidad final es:

$$\pi(\theta|y) = \pi(\theta|s) = \frac{\left(\frac{1}{\theta^2} + \frac{1}{2-e^\theta - e^{-\theta}} \right)^{1/2} \left(\frac{\theta}{e^\theta - 1} \right)^n e^{\theta s} h_n(s)}{h_n(s) \int_{-\infty}^{\infty} \left(\frac{1}{\theta^2} + \frac{1}{2-e^\theta - e^{-\theta}} \right)^{1/2} \frac{e^n e^{\theta s}}{(e^\theta - 1)^n} d\theta}$$

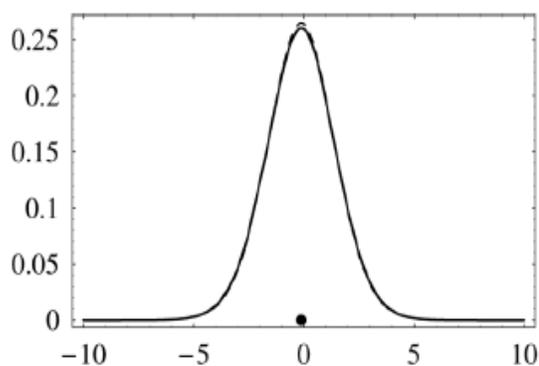
Como $\int_{-\infty}^{\infty} \left(\frac{1}{\theta^2} + \frac{1}{2-e^\theta - e^{-\theta}} \right)^{1/2} \frac{e^n e^{\theta s}}{(e^\theta - 1)^n} d\theta = \text{constante}$, se tiene:

$$\pi(\theta|s) \propto \left(\frac{1}{\theta^2} + \frac{1}{2-e^\theta - e^{-\theta}} \right)^{1/2} \frac{\theta^n e^{\theta s}}{(e^\theta - 1)^n} \quad (23)$$

Se puede ver que las distribuciones fiducial y posterior de θ , no son iguales; pero existe una gran similitud en la práctica. Se muestran los casos que O'Reilly y Rueda (2007) graficaron para muestras particulares de la exponencial truncada para algunos valores del parámetro θ , el tamaño de la muestra y el estimador máximo verosímil del parámetro¹⁷.

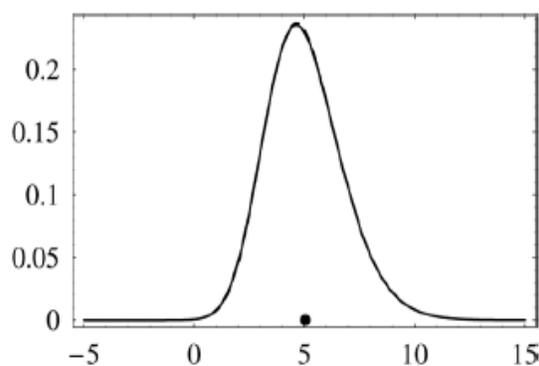
En las gráficas la línea sólida es la densidad fiducial y la línea segmentada representa a la densidad final (Gráficas 3.7., 3.8., 3.9., 3.10., 3.11., 3.12. 3.13. y 3.14.).

Gráfica 3.7



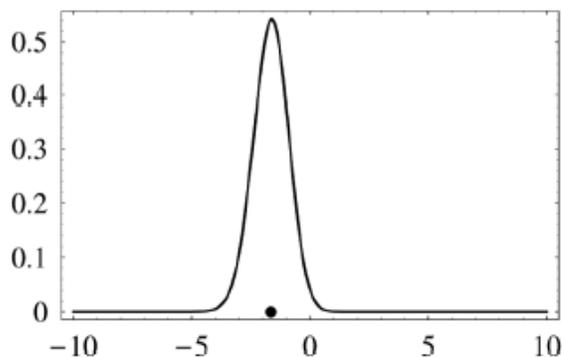
$$(\theta, n, \hat{\theta}, s) = (1, 5, -0.13, 2.44)$$

Gráfica 3.8



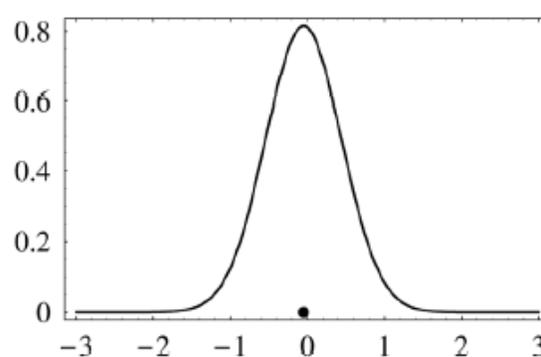
$$(\theta, n, \hat{\theta}, s) = (2, 10, 5.04, 8.08)$$

Gráfica 3.9



$$(\theta, n, \hat{\theta}, s) = (-2, 25, -1.66, 9.19)$$

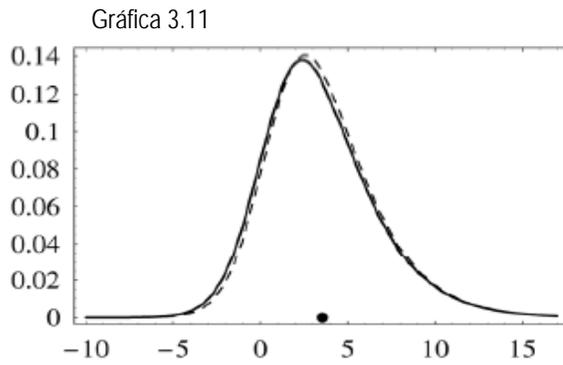
Gráfica 3.10



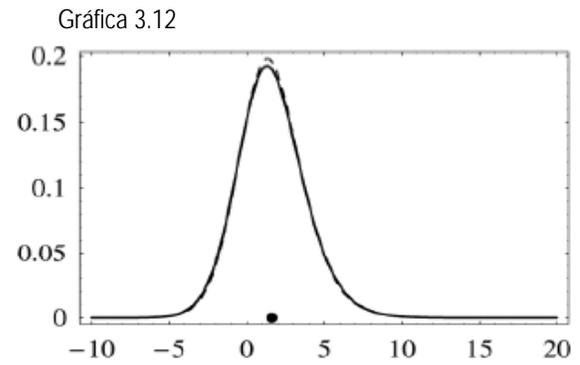
$$(\theta, n, \hat{\theta}, s) = (0, 50, -0.05, 24.79)$$

¹⁷ Los autores O'Reilly y Rueda calculan este estimador resolviendo numéricamente la ecuación

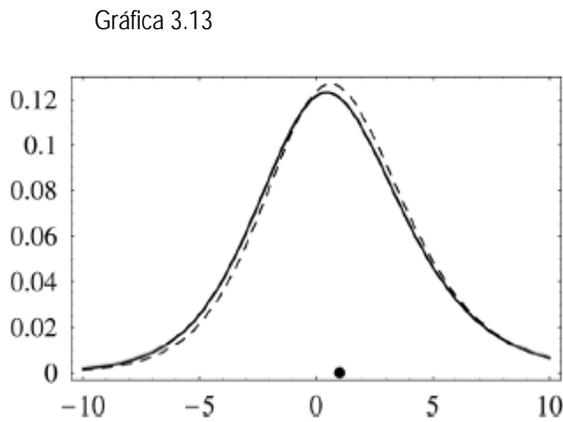
$$\frac{1}{\theta} - \frac{1}{1 - e^{-\theta}} + \frac{s}{n} = 0.$$



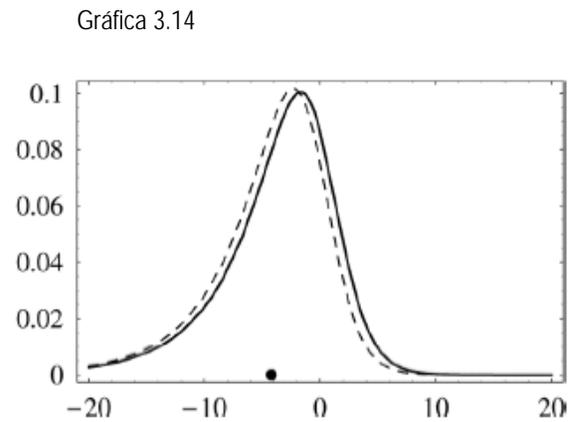
$$(\theta, n, \hat{\theta}, s) = (-4, 2, 3.54, 1.50)$$



$$(\theta, n, \hat{\theta}, s) = (0, 3, 1.61, 1.88)$$



$$(\theta, n, \hat{\theta}, s) = (-2, 1, 0.01, 0.58)$$



$$(\theta, n, \hat{\theta}, s) = (-3, 1, -4.19, 0.22)$$

Ejemplo 3.4. Ji cuadrada no central

Sea X_1, \dots, X_n v.a. iid $N(\mu_i; 1)$, $i = 1, \dots, n$, y se quiere obtener la distribución fiducial

para $\sum_{i=1}^n \mu_i^2$.

Desde el punto de vista fiducial, de acuerdo a Stein (1959) μ_1, \dots, μ_n son independientes

e idénticamente distribuidas $N(X_i; 1)$, $i = 1, \dots, n$.

Por otro lado la estadística suficiente minimal para $\sum_{i=1}^n \mu_i^2 = \theta$ es $\sum_{i=1}^n X_i^2$ y considerando el siguiente teorema.

Teorema 2¹⁸.

Si X_1, \dots, X_n son v. a. ii con medias μ_1, \dots, μ_n , respectivamente y con varianza igual a la unidad, entonces $S = \sum_{i=1}^n X_i^2$ se distribuye como una Ji cuadrada no central con n grados

de libertad y $\lambda = \frac{1}{2} \sum_{i=1}^n \mu_i^2$ como parámetro de centralidad y la función es¹⁹:

$$g_0(s; n, \lambda) = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i s^{\frac{1}{2}(n+2i)-1} e^{-\frac{s}{2}}}{i! 2^{\frac{1}{2}(n+2i)} \Gamma\left(\frac{n+2i}{2}\right)}; \quad 0 \leq s < \infty$$

Calculando la función de distribución de la estadística suficiente minimal, queda:

$$G_0(s; n, \lambda) = \int_0^s e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i u^{\frac{1}{2}(n+2i)-1} e^{-\frac{u}{2}}}{i! 2^{\frac{1}{2}(n+2i)} \Gamma\left(\frac{n+2i}{2}\right)} du, \text{ considerando que } \sum_{i=1}^n \mu_i^2 = \theta, \text{ entonces}$$

¹⁸ Ver demostración en Graybill, F. "An introduction to linear statistical models" Volumen 1.

¹⁹ La notación para la distribución Ji-cuadrada no central con grados de libertad n y parámetro de centralidad λ es $\chi_{(n, \lambda)}^2$.

$$G_0(s; n, \theta) = \int_0^s e^{-\frac{\theta}{2}} \sum_{i=0}^{\infty} \frac{\left(\frac{\theta}{2}\right)^i u^{\frac{1}{2}(n+2i)-1} e^{-\frac{u}{2}}}{i! 2^{\frac{1}{2}(n+2i)} \Gamma\left(\frac{n+2i}{2}\right)} du$$

$$G_0(s; n, \theta) = \int_0^s e^{-\frac{\theta}{2}} \sum_{i=0}^{\infty} \frac{\theta^i u^{\frac{1}{2}(n+2i)-1} e^{-\frac{u}{2}}}{i! 2^{\frac{1}{2}(n+2i)+i} \Gamma\left(\frac{n+2i}{2}\right)} du \quad (24)$$

Cuando ya se observó la muestra la distribución fiducial, de acuerdo a lo visto en este trabajo, queda como:

$$F_{fid}(\theta^*) = 1 - G_0(s_0; n, \theta^*)$$

$$= 1 - \int_0^{s_0} e^{-\frac{\theta}{2}} \sum_{i=0}^{\infty} \frac{\theta^i u^{\frac{1}{2}(n+2i)-1} e^{-\frac{u}{2}}}{i! 2^{\frac{1}{2}(n+2i)+i} \Gamma\left(\frac{n+2i}{2}\right)} du \quad (25)$$

Para revisar si cumple el criterio de contrastabilidad monótona para la distribución fiducial, se tiene que tiene que ser $F_{fid}(\theta^*)$ decreciente como función θ .

Considerando (25) como $F_{fid}(\theta^*) = 1 - e^{-\frac{\theta}{2}} \sum_{i=0}^{\infty} \frac{\theta^i}{i! 2^i} \int_0^{s_0} \frac{u^{\frac{1}{2}(n+2i)-1} e^{-\frac{u}{2}}}{2^{\frac{1}{2}(n+2i)} \Gamma\left(\frac{n+2i}{2}\right)} du$, donde la integral

indica la función de distribución acumulada de una gama $\left(\alpha = \frac{n+2i}{2}, \beta = 2\right)$ denotándola

como $R\left(s_0; \alpha = \frac{n+2i}{2}, \beta = 2\right)$, entonces

$$F_{fid}(\theta^*) = 1 - e^{-\frac{\theta}{2}} \sum_{i=0}^{\infty} \frac{\theta^i}{i!2^i} R\left(s_0; \alpha = \frac{n+2i}{2}, \beta = 2\right).$$

Considerando $\theta_1 < \theta_2$, se tiene

$$1 - e^{-\frac{\theta_1}{2}} \sum_{i=0}^{\infty} \frac{\theta_1^i}{i!2^i} R\left(s_0; \alpha = \frac{n+2i}{2}, \beta = 2\right) > 1 - e^{-\frac{\theta_2}{2}} \sum_{i=0}^{\infty} \frac{\theta_2^i}{i!2^i} R\left(s_0; \alpha = \frac{n+2i}{2}, \beta = 2\right),$$

ya que $\frac{R\left(s_0; \alpha = \frac{n+2i}{2}, \beta = 2\right)}{i!2^i}$ siempre es positiva; por lo que $F_{fid}(\theta^*)$ es decreciente para toda θ .

Para revisar si existe certidumbre en la contrastabilidad extrema se tiene que:

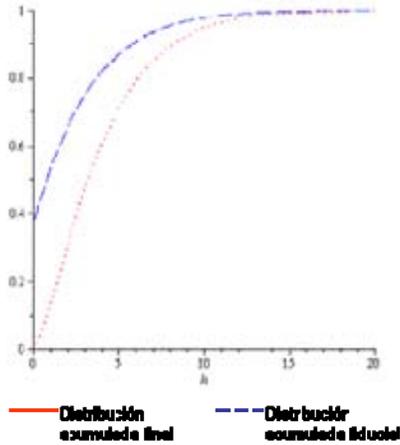
$$\begin{aligned} \lim_{\theta \rightarrow 0} F_{fid}(\theta^*) &= \lim_{\theta \rightarrow 0} \left(1 - e^{-\frac{\theta}{2}} \sum_{i=0}^{\infty} \frac{\theta^i}{i!2^i} R\left(s_0; \alpha = \frac{n+2i}{2}, \beta = 2\right) \right) \\ &= 1 - \chi_{(n)}^2 \end{aligned}$$

Como puede observarse no produjo solamente uno, por lo que hay una masa en uno de sus extremos, siendo la distribución fiducial "defectuosa", ya que no integra a uno, aunque la función pertenezca a la familia exponencial.

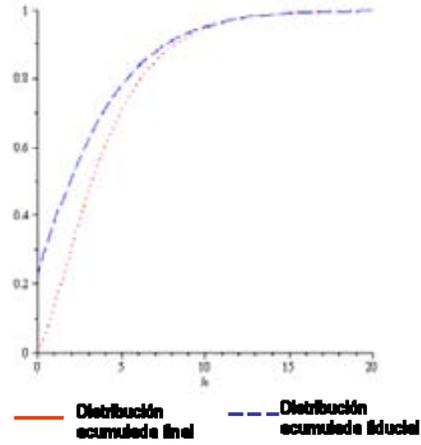
En las gráficas 3.15 y 3.16 se muestran las funciones de distribución acumuladas final y fiducial y puede apreciarse como la fiducial no comienza en cero, sino que acumula un valor en ese punto. La gráfica 3.15 da una acumulación en cero mayor que la gráfica 3.16 debido a la estimación de la estadística suficiente minimal. En ambas gráficas puede

apreciarse que a medida que θ tiende a ser grande las dos funciones de distribución acumulada son materialmente iguales.

Gráfica 3.15



Gráfica 3.16



Obteniendo ahora la distribución fiducial se deriva la expresión (25):

$$\begin{aligned}
 f_{fid}(\theta^*) &= -\frac{\partial G_0(s_0; n, \theta^*)}{\partial \theta^*} \\
 &= -\frac{\partial}{\partial \theta^*} \left(\int_0^{s_0} e^{-\frac{\theta^*}{2}} \sum_{i=0}^{\infty} \frac{(\theta^*)^i u^{\frac{1}{2}(n+2i)-1} e^{-\frac{u}{2}}}{i! 2^{\frac{1}{2}(n+2i)+i} \Gamma\left(\frac{n+2i}{2}\right)} du \right) \\
 &= \int_0^{s_0} e^{-\frac{\theta^*}{2}} \sum_{i=0}^{\infty} \frac{(\theta^*)^i u^{\frac{1}{2}(n+2i)-1} e^{-\frac{u}{2}}}{i! 2^{\frac{1}{2}(n+2i)+i} \Gamma\left(\frac{n+2i}{2}\right)} \left(\frac{1}{2} - \frac{i}{\theta}\right) du \\
 &= \int_0^{s_0} \sum_{i=0}^{\infty} \frac{e^{-\frac{\theta^*}{2}} \left(\frac{\theta^*}{2}\right)^i}{i!} \left(\frac{1}{2} - \frac{i}{\theta}\right) \left(\frac{u^{\frac{1}{2}(n+2i)-1} e^{-\frac{u}{2}}}{2^{\frac{1}{2}(n+2i)} \Gamma\left(\frac{n+2i}{2}\right)} \right) du \\
 &= \int_0^{s_0} \sum_{i=0}^{\infty} \frac{e^{-\frac{\theta^*}{2}} \left(\frac{\theta^*}{2}\right)^i}{i!} \left(\frac{1}{2} - \frac{i}{\theta}\right) m_{\left(\frac{n+2i}{2}\right)}(u) du
 \end{aligned}$$

$$f_{fid}(\theta^*) = \sum_{i=0}^{\infty} \frac{e^{-\frac{\theta^*}{2}} \left(\frac{\theta^*}{2}\right)^i}{i!} \left(\frac{1}{2} - \frac{i}{\theta}\right) \int_0^{s_0} m_{\left(\frac{n+2i}{2}\right)}(u) du \quad (26)$$

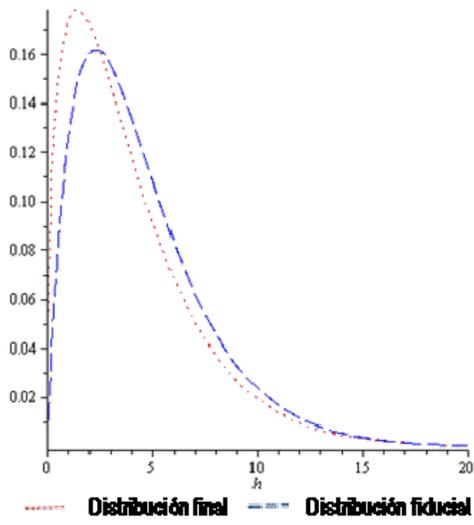
Donde $m_{\left(\frac{n+2i}{2}\right)}(u)$ = densidad de una Ji-cuadrada central, por lo que $\int_0^{s_0} m_{\left(\frac{n+2i}{2}\right)}(u) du = M\left(s_0; \frac{n+2i}{2}\right)$ es la distribución acumulada de esa Ji-cuadrada central con $\frac{n+2i}{2}$ grados de libertad. Por lo tanto la densidad fiducial de θ es la suma de un infinito número de distribuciones acumuladas Ji-cuadradas centrales ponderadas por dos factores, uno la densidad Poisson y el otro $\left(\frac{1}{2} - \frac{i}{\theta}\right)$.

La distribución posterior para este ejemplo es dada en Stein (1959) para $\sum_{i=1}^n \mu_i^2$, la cual es una Ji- cuadrada no central con n grados de libertad y parámetro de no-centralidad $S = \sum_{i=1}^n x_i^2$.

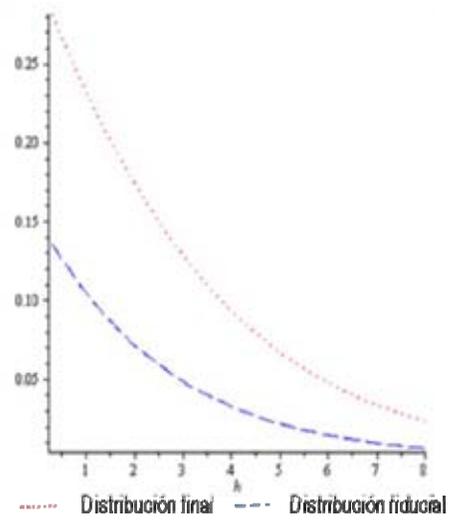
$$\pi(\theta/s) = e^s \sum_{j=0}^{\infty} \frac{(s)^j \theta^{\frac{1}{2}(n+2j)-1} e^{-\frac{\theta}{2}}}{j! 2^{\frac{1}{2}(n+2j)} \Gamma\left(\frac{n+2j}{2}\right)}; \quad 0 \leq \theta < \infty \quad (27)$$

Es claro que las expresiones fiducial (26) y Bayesiana (27) son diferentes, pero esto no impide que se parezcan numéricamente, como puede apreciarse en las gráficas se parecen mucho cuando la θ es grande, pero difieren y se muestra la masa cuando θ tiende a cero.

Gráfica 3.17



Gráfica 3.18



Conclusiones

Este trabajo da un recorrido primeramente por el método clásico del cálculo de intervalos de confianza, a través del uso de cantidades pivotaes, revisando la más común que es la cantidad pivotal exacta, pero como no es siempre fácil proponerla para algunos problemas, se requiere encontrar alternativas que permitan calcular los intervalos, por lo que se señala el uso de cantidades pivotaes asintóticas, mostrando un resumen que Sprott dio sobre ellas.

Desde que Fisher (1930) expuso su argumento fiducial, siempre ha sido controversial, pero dio una alternativa importante para la construcción de intervalos. Así en este sentido se mostraron también los argumentos fiducial y fiducial implícito para el cálculo de los intervalos, que ahora bajo este enfoque reciben el nombre de intervalos fiduciales y se señaló que cuando las distribuciones vienen de familias de grupo existe coincidencia con estos últimos intervalos y los de confianza calculados con las cantidades pivotaes. También se revisó el procedimiento Bayesiano para la construcción de intervalos de probabilidad (*a-posteriori*) de densidad máxima, señalando que se da una coincidencia con los intervalos fiduciales cuando en los primeros se da una distribución inicial (*a-priori*) del parámetro θ , no informativa, y el modelo es de grupo.

Al final de las secciones fiducial y Bayesiana, se expuso el ejemplo de estimar el parámetro por un intervalo fiducial y de probabilidad de densidad máxima de la función Bernoulli, donde resultó que la distribución fiducial no coincide con la distribución de la cantidad pivotal aproximada usada en la forma clásica y que tampoco concordó con la

distribución final, cuando se dio una distribución *a-priori* de p no informativa como

$\pi(p) = p^{-\frac{1}{2}}(1-p)^{-\frac{1}{2}}$, pero numéricamente, se hizo evidente, que difieren muy poco.

Esta característica de diferencia algebraica pero de gran semejanza numérica se mostró en el capítulo 3. A través de tres ejemplos se revisaron las condiciones de Lindley sobre la coincidencia de las distribuciones fiducial y final, mostrando con dos de ellos que al dar parámetros que no son de localización efectivamente no se tendrá igual forma algebraica de las dos distribuciones, pero al graficarlas, el parecido es notable y con el último ejemplo se reveló una función fiducial que mostró en el valor cero del parámetro una acumulación, difiriendo completamente de la distribución final, pero ello no impidió que su aproximación numérica fuera buena.

Se puede concluir que los argumentos fiducial y Bayesiano seguirán teniendo críticas, pero fue interesante mostrar que a pesar de todo, son enfoques que en ciertas condiciones coinciden entre sí y con el método clásico; pero que además cuando esas condiciones no se dan, la aproximación numérica es excelente; esta coincidencia muestra algo importante en el sentido de que el procedimiento fiducial debe tener "algo" que le permite generar aproximaciones excelentes.

Por estas aproximaciones me permito decir que vale la pena intentar un enfoque, como es el fiducial, para hacer inferencias del parámetro con la cualidad de dar una distribución de probabilidad al parámetro a estimar y conservar el hecho de que la muestra al ser observada se vuelve fija, crítica que ha sido avasallante en la estadística clásica.

Bibliografía

1. Andersen, E. B (1970). "Sufficiency and Exponential Families for Discrete Sample Spaces". *Journal of the American Statistical Association*, Vol. 65, No. 331, 1248-1255.
2. Aznar, G. A. (1984). "Buscando el modelo econométrico útil". *Estadística Española* no. 102, 5 - 11.
3. Barnard, G. A. (1963a). "Some logical aspects of the Fiducial Argument". *Journal of the Royal Statistical Society*, B 25, 114-118.
4. Barnard, G. A. (1963b). "Some logical aspects of the Fiducial Argument". *Bulletin of the International Statistical Institute*. 40(2), 870-883.
5. Barnard, G. A. (1987). "R. A. Fisher – a true Bayesian?" *International Statistical Review*. 55, 2, 183-189.
6. Bin Dong, L. (2004). "The Behrens-Fisher Problem: An empirical likelihood approach". *Econometrics Working Paper* EWP0404. Department of Economics, University of Victoria, B. C. Canada.
7. Bolduc, D; Khalaf, L y Yélou, C. (2005). "Identification robust confidence sets on parameters ratio with application to discrete choice models". *Journal of Economic Literature*. C 10. C35. R 40.

8. Bolstad, W. M. (2007). *Introduction to Bayesian Statistics*. Second Edition. Wiley-Interscience. John Wiley & Sons. USA.
9. Box, G. y Tiao, G. (1992). *Bayesian Inference in Statistical Analysis*. John Wiley and Sons, Inc. NY. USA.
10. Cox, D. R. y Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman and Hall. First edition. London.
11. Chen, M_H; Shao, Q.M y Ibrahim, J. (2000). *Montecarlo Methods in Bayesian Computation*. Springer. NY. USA.
12. Engen, S. y Lillegard, M. (1997). "Stochastic Simulations Conditioned on Sufficient Statistics". *Biometrika*. Vol. 84, No. 1, 235-240.
13. Feller, W. (1985). *Introducción a la Teoría de Probabilidades y sus aplicaciones*. Editorial Limusa, segunda edición. México.
14. Fieller, E. C. (1954). Some problems in interval estimations. Royal Statistical Society. Series B, Vol. 16. No. 2. 175 – 185.
15. Fisher, R. A. (1959). *Statistical Methods and Inference*. Second Edition, London, Oliver and Boyd.
16. Fraser, D. A. S. (1961a). "On fiducial inference". *The Annals of Mathematical Statistics*, 32, 661-676.

17. Fraser, D. A. S. (1961b). "The Fiducial Method and Invariance". *Biometrika*. 48, 261-281.
18. Gelman, A.; Carlin, J. B.; Stern, H. S. and Rubin, D. (2004). *Bayesian Data Analysis*. Chapman & Hall/CRC. Second edition. New York, USA.
19. Graybill, F. A. (1961). *An Introduction to Linear Statistical Models*, Volume I. McGraw-Hill. USA.
20. Hirschman, I. I. (1962). *Infinite Series*. Holt, Rinehart and Winston. New York.
21. Hogg, R. V. y Craig. A. T. (1978). *Introduction to Mathematical Statistics*. Macmillan Publishing Co. Fourth edition. New York.
22. Knight, K. (2000). *Mathematical Statistics*. Chapman & Hall/CRC. USA.
23. Lange, K. (1998). *Numerical Analysis for Statisticians*. Springer. New York.
24. Lee, P. M. (1989). *Bayesian Statistics*. Oxford University Press. New York.
25. Lehmann, E. L. (1998). *Elements of Large-Sample Theory*. Springer text in Statistics. USA.
26. Lehmann, E. L. Casella, G. (1999). *Theory of point Estimation*. Springer Text in Statistics. Second Edition. NY. USA.

27. Lew, J. S. (1975). "On Some Relations between the Laplace and Mullin Transforms". *IBM. Journal of Research and Development*. 5682-586.
28. Lindley, D. V. (1958). "Fiducial Distribution and Baye's Theorem". *Journal of the Royal Statistical Society, B* 20, 102-107.
29. Lindley, D. V. (1961). "The use of Prior Probability Distributions in Statistical Inference and Decisions". *Proc. Fourth Berkeley Symp.* 1, 453-468.
30. Lindley, D. V. (1980) *Introduction to Probability & Statistics. From a Bayesian viewpoint. Part 1. Probability*. Cambridge University Press. First paperback edition. NY. USA.
31. Lindley, D. V. (1980) *Introduction to Probability & Statistics. From a Bayesian viewpoint. Part 2. Inference*. Cambridge University Press. First paperback edition. NY. USA.
32. Lindsey, J. K. (1996). *Parametric Statistical Inference*. Clarendon Press Oxford. Great Britain.
33. Lingren, B. (1993). *Statistical Theory*. Chapman & Hall. New York, USA.
34. Lomnicki, Z. A. (1967). "On the Distribution of Products of Random Variables". *Journal of the Royal Statistical Society. B*, Vol. 29, No. 3. 513-524.
35. Marsden J. (1973). *Basic Complex Analysis*. W. H. Freeman and Company. San Francisco Ca. USA.

36. Mood, A.; Graybill, F. y Boes, D. (1985). *Introduction to the Theory of Statistics*. International Student Edition. Third Edition.
37. O'Reilly T. F. y Rueda, R. (1992). "Goodness of fit for inverse Gaussian distribution". *Canadian Journal of Statistics*, 20, 387-397.
38. O'Reilly T. F. (2003). "Significance Distributions". Comunicaciones Técnicas IIMAS. Preimpreso no. 117.
39. O'Reilly T. F. y Rueda, R. (2007). "Inferences in truncated exponential distributions". *Communications in Statistics—Theory and Methods*, 36 2207–2212.
40. Ortega, I. F. J. y Basulto, S. J. (2003). "Distribuciones a priori unidimensionales en Modelos No Regulares: Medidas de Información". *Estadística Española*. Vol. 45, Núm. 154, 363-383.
41. Pederzen, J. G. (1978). "Fiducial Inference". *International Statistical Review*. 46, 147-170.
42. Ruben, H. (2002). "A simple conservative and robust solution of the Behrens-Fisher problem". *Sankhy-a : The Indian Journal of Statistics*, Volume 64, Series A, Pt.1, pp. 139-155.
43. Rueda, R. (2007). "¿Fisher estaba en lo correcto?" *Jornadas de Estadística*. IIMAS, UNAM.

44. Savage, L. (SF). "The Foundations of Statistics Reconsidered". *Fourth Berkeley Symposium University of Michigan*. 575-586.
45. Shanbhag, D. N (1970). "The Characterizations for Exponential and Geometric Distributions". *Journal of the American Statistical Association*, Vol. 65, No. 331, 1256-1259.
46. Singh, R. S. (2001). "Bayes and empirical Bayes procedures for selecting good populations from a Translated Exponential Family". *Empirical Bayes and Likelihood Inference*. Springer. NY. USA.
47. Sondow, J. (1994). "Analytic Continuation of Riemann's Zeta Function and Values at Negative integers Via Euler's Transformation of Series". *American Mathematical Society*. Vol. 120, No. 2, 421-424.
48. Sondow, J. (1998). "An ant symmetric formula for Euler's constant". *Mathematics Magazine* 71. 219-220.
49. Sondow, J. (1998). "Tue Riemann Hypothesis, Simple Zeros and the Asymptotic Convergence Degree of Improper Riemann Sums". *American Mathematical Society*. Vol 126, No. 5, 1311-1314.
50. Spiegel, M. R. (1964). *Theory and problems of Complex Variables with an introduction to conformal mapping and its applications*. Schaum Publishing Co. New York.

51. Sprott, D. A. (1963). "A transformation Model for the Investigation of Fiducial Distributions". *Bulletin of the International Statistical Institute*. 40, 2, 856-869.
52. Sprott, D. A. (1966). "Statistical Estimation – Some Approaches and Controversies". *Statistische Hefte* 6, 97-111.
53. Sprott, D. A. y Kalbfleisch, J. (1969). "Examples of likelihoods and comparisons with point estimates and large sample approximations". *American Statistical Association Journal*. 468-484.
54. Sprott, D. A. (1973). "Normal likelihoods and their relation to large sample theory of estimation". *Biometrika*, 60, 457-465.
55. Sprott, D. A. (1975). "Application of Maximum Likelihood to Finite Samples". *Sankhya Ser. B* 37, 259-270.
56. Sprott, D. A. y Kalbfleisch, J. G. (1967). "Fiducial Probability". *Statistische Hefte* 2, 99-109.
57. Sprott, D. A. y Kalbfleisch, J. G. (1969). "Examples of Likelihoods and Comparisons with Point Estimates and Large Sample Approximations". *Journal American Statistical Association*, 64, 468-484.
58. Sprott, D. A. y Viveros, R. (1984). "The Interpretation of Maximum Likelihood Estimation". *Canadian Journal of Statistics*, 12, 27-38.

59. Stein, C. (1959). "An Example of Wide Discrepancy between Fiducial and Confidence Intervals". *The Annals of Mathematical Statistics*, Vol. 30, No. 4, 877-880.
60. Tukey, J. W. (1957). "Some Examples with Fiducial Relevance". *The Annals of Mathematical Statistics*, 28, 687-695.
61. Tsui, K-W, Tang, S. (2005). Distributional Property of the generalized p-value for the Behrens-Fisher Problem with Applications to Multiple Testing. *University of Wisconsin, Technical Report No. 1111*.
62. Welch, B. L. y Peers, H. W (1963). On formulae for Confidence points Based on Integrals of Weighted Likelihood. *Journal of the Royal Statistical Society*, B 25, 318.
63. Yin, M. y Ghosh M. (2001). *Bayesian and likelihood inference for the generalized Fieller- Creasy problem. Empirical Bayes and Likelihood Inference*. Springer. NY. USA.