

SUBJETIVIDAD, OBJETIVIDAD Y COHERENCIA EN EL ANÁLISIS DE LA EXPERIENCIA

Pere Agell
Rafael Andreu
Josep M. Rosanas

SUBJETIVIDAD, OBJETIVIDAD Y COHERENCIA EN EL ANÁLISIS DE LA EXPERIENCIA

Pere Agell¹

Rafael Andreu²

Josep M. Rosanas³

Resumen

Este trabajo aborda el problema de cómo acumular información procedente de datos obtenidos de la observación de un determinado fenómeno, para predecir su comportamiento coherentemente con aquella información. Se parte de una hipótesis sobre las características estructurales básicas del fenómeno de interés (por ejemplo, que los datos observados proceden de un proceso de Bernoulli). Además de este conocimiento, se necesita información del parámetro(s) que gobierna(n) dicho proceso.

Con objeto de derivar un procedimiento general que aúne un punto de vista inicial subjetivo y otro procedente de la observación, se explica de qué forma proceder, para después derivar un proceso de acumulación de datos basado en realizar sucesivos experimentos (como muestreos) y, aplicando el teorema de Bayes, obtener sucesivas distribuciones a posteriori de los parámetros. Procediendo así puede demostrarse general y rigurosamente que las sucesivas distribuciones a posteriori no pueden ser cualesquiera, sino que deben pertenecer a una familia que depende del tipo de fenómeno subyacente que se supone «genera» los datos que se van observando.

Un resultado general adicional es que, si alternativamente se agruparan los datos de los experimentos y se procediera como si se hubieran generado en uno sólo, se obtendría una distribución a posteriori no sólo de la misma familia, sino exactamente coincidente con la obtenida secuencialmente. En otras palabras: a partir de la absoluta ignorancia original, llegamos siempre al mismo resultado.

Con este bagaje puede replantearse el proceso partiendo de una información subjetiva obtenida no con base en experimentos formales, sino en el conocimiento directo de la realidad. Para ser coherente con lo anterior, la distribución de probabilidad mediante la cual se exprese dicho conocimiento puede ser subjetiva pero no cualquiera: debe pertenecer a la familia de distribuciones asociada al tipo de fenómeno subyacente que se esté considerando y, en consecuencia, corresponder a una muestra o experimento bien definido.

Clasificación JEL: B1, B2, B3.

¹ Profesor Emérito Análisis de Decisiones, IESE Business School.

² Profesor Sistemas de Información, IESE Business School.

³ Profesor Contabilidad y Control, IESE Business School.

SUBJETIVIDAD, OBJETIVIDAD Y COHERENCIA EN EL ANÁLISIS DE LA EXPERIENCIA

Introducción

En un reciente artículo en el *New York Times*, revisando un libro de Sharon McGrayne sobre la historia del teorema de Bayes, el conocido matemático John Allen Paulos decía que *«the nub of the differences between them [es decir, entre los estadísticos bayesianos ortodoxos y los llamados “frecuencionalistas”] is that for Bayesians the prior can be a subjective expression of the degree of belief in a hypothesis, even one about a unique event or one that has yet never occurred. For frequentists the prior must have a more objective foundation; ideally is the relative frequency of events in repeatable, well-defined experiments»*.

Los bayesianos han sido tradicionalmente criticados por esta razón, a partir de la creencia de que la objetividad es algo que una ciencia debe poseer ineluctablemente, mientras que la subjetividad, en cambio, proporciona bases poco sólidas sobre las que construir el conocimiento. Es decir, se considera que la subjetividad es, hasta cierto punto, arbitraria.

El razonamiento bayesiano, que refleja la cita inicial en el libro de Sharon McGrayne de Keynes (*«When the facts change, I change my opinion. What do you do, sir?»*), es, sin embargo, impecable. Uno empieza con un cierto conocimiento de un fenómeno y, conforme va viendo lo que ocurre en la realidad, cambia sus opiniones sobre el mismo. El único problema está, pues, en el principio: ¿De dónde sale una «posición inicial» respecto a un fenómeno que se desea estudiar (una «distribución *a priori*» en el ámbito de la estadística; a veces, simplemente llamada «prior») antes de realizar ningún experimento formal?

Datos, subjetividad y experiencia

Curiosamente, aunque los datos y su análisis estadístico son útiles para tomar decisiones, en el contexto empresarial la subjetividad es un elemento indispensable para tomarlas: precisamente por ello, hay empresas que tienen éxito y otras que no. Empresas como Hewlett-Packard, Apple, Microsoft y muchas otras que han llegado a ser grandes (si no todas) han empezado con una creencia subjetiva por parte del fundador o grupo de fundadores sobre un producto o una manera de hacer las cosas que podía tener éxito. Y la idea no era compartida por el resto de la humanidad (si lo hubiera sido, otras personas hubieran hecho lo mismo y la idea habría tenido menos éxito).

Por supuesto, los éxitos que todos conocemos han coexistido con un buen número de fracasos también basados en ideas subjetivas sobre qué hacer y cómo hacerlo. En cualquier caso, quien posee (o cree poseer) una información «subjetiva» la usa para tomar sus decisiones. Si está bien construida, de acuerdo con criterios matemáticos de coherencia, será mejor usarla que descartarla: o bien el tomador de decisiones sabe lo que cree que sabe o no lo sabe; si no, es simplemente cuestión de reconocerlo explícita y coherentemente. La pregunta es cómo.

¿Podemos saber algo que proceda directamente de la experiencia, sin haber recogido datos en ningún experimento «formal»? Parece indudable que sí, a juzgar por lo visto en los ejemplos citados y en muchos otros, y es razonable que tomemos decisiones basadas en este conocimiento. Sería ridículo despreciarlo por el hecho de ser subjetivo.

Pero hacerlo requiere ser cuidadosos para evitar incoherencias, porque nuestra mente a veces nos juega malas pasadas. Quien más quien menos, en alguna ocasión se ha encontrado con creencias inconsistentes entre sí. Fischhoff, Slovic y Lichtenstein (1988) proporcionan un interesante listado de los estados de ánimo asociados con tener creencias incoherentes y con sus posibles consecuencias: no percatarse, percatarse, vivir en la incoherencia o tratar de formarse una opinión coherente. Si queremos comportarnos racionalmente (aunque sea de manera limitada, como repetidamente ha señalado Herbert Simon), deberíamos tratar de eliminar las incoherencias que dichas limitaciones pueden provocar.

En el contexto que nos ocupa, esto implica un escrupuloso examen de nuestras probabilidades subjetivas para evitar que resulten inconsistentes en sí mismas (lo que ocurre con cierta frecuencia) o con el fenómeno objeto de estudio.

El problema básico consiste en cómo acumular la información procedente de datos obtenidos de la observación de un determinado fenómeno para predecir su comportamiento cada vez mejor. Normalmente, se acepta partir de una hipótesis acerca de las características estructurales básicas del fenómeno de interés, que se supone de un determinado tipo (por ejemplo, podemos suponer que los datos que observamos proceden de un proceso probabilístico de Bernoulli). Además de este conocimiento, necesitaremos, para trabajar, alguna información respecto al parámetro (o parámetros) que gobierna(n) dicho proceso.

Con objeto de derivar un procedimiento general, que aúne un punto de vista inicial subjetivo y otro procedente de la observación de determinados datos, procederemos como sigue: si partimos de una situación en la que no disponemos de datos ni información alguna de donde sacar una distribución inicial de dichos parámetros, lo coherente es admitir que, en esa situación, nuestra distribución de probabilidad inicial de dicho(s) parámetro(s) (denominada «distribución *a priori* inicial») es independiente del valor real (y absolutamente desconocido) de los mismos.

Esta consideración de partida permite iniciar un proceso de acumulación de datos a base de realizar sucesivos experimentos (muestreos, por ejemplo) y, aplicando el teorema de Bayes, obtener sucesivas distribuciones *a posteriori* de los parámetros, que van incorporando toda la información obtenida a través de dichos datos. Procediendo así puede demostrarse general y rigurosamente que las sucesivas distribuciones *a posteriori* no pueden ser cualesquiera, sino que deben pertenecer a una familia determinada que depende del tipo de fenómeno subyacente que se supone «genera» los datos que vamos observando.

Un resultado general adicional es que, si alternativamente agrupamos los datos de los sucesivos experimentos y procedemos como si se hubieran generado en uno sólo, obtendríamos una distribución *a posteriori* no sólo de la misma familia, sino exactamente coincidente con la

obtenida secuencialmente. En otras palabras, mientras mantengamos las hipótesis de que el fenómeno subyacente es el mismo y que el valor real de sus parámetros no varíe, sucesivas muestras y cálculos de distribuciones *a priori/posteriori* son equivalentes a cualquier otra sucesión de muestras que en conjunto contengan los mismos datos, incluida la única que las reúne todas: a partir de la absoluta ignorancia original, llegamos siempre al mismo resultado.

Con este bagaje puede entonces replantearse el proceso partiendo de una información subjetiva obtenida no con base en experimentos formales, sino en nuestro conocimiento directo de la realidad. Para ser coherente con lo anterior, la distribución de probabilidad mediante la cual expresemos dicho conocimiento puede ser subjetiva pero no cualquiera: debe pertenecer a la familia de distribuciones asociada al tipo de fenómeno subyacente que estemos considerando (y así corresponder implícitamente a una muestra o experimento equivalente). Si no, lo que creemos saber no sería consistente con la hipótesis del tipo de fenómeno subyacente supuesto. Esto pone de manifiesto, rigurosa y consistentemente, que la elección de un tipo de fenómeno determinado forma parte de nuestro conocimiento inicial, que debemos preservar si no lo cambiamos explícitamente (por ejemplo, porque sucesivos datos nos hacen pensar que es otro, con el que deberíamos entonces reinterpretar los datos procedentes de los experimentos).

Visto de otra manera: por mucho que una persona pueda hoy «saber algo» con respecto a un determinado fenómeno, siempre ha habido algún momento antes en que no ha sabido nada; y, por tanto, mirando atrás, lo que sabe es equivalente a un experimento que podemos calcular. Es decir, nuestro conocimiento subjetivo no puede (ni debe) ser arbitrario: la consistencia nos exige expresarlo a través de una familia de distribuciones determinada por el modelo asociado al tipo de fenómeno supuesto.

A partir de aquí podemos interpretar que lo que sabemos en términos intuitivos-subjetivos es equivalente a un experimento «calculado» con base en la función de distribución de la familia que corresponda; y usar los datos implícitos en ese experimento equivalente para añadirlos a los que podamos ir obteniendo de nuevos experimentos. Así, entendiendo nuestra información subjetiva como equivalente a una colección de datos, podemos usar la *a priori* «subjetiva» para calcular por Bayes la *a posteriori*, y el resultado debe ser, necesaria y coherentemente, el mismo.

En consecuencia, resulta absurdo separar «ideológicamente» a los expertos en estadística entre «bayesianos» y «no bayesianos»: ambos deben utilizar los mismos métodos para calcular sus distribuciones *a posteriori* y la única diferencia reside en si decidimos usar o no la información intuitiva-subjetiva, que puede siempre demostrarse equivalente a los resultados de un experimento determinado y calculable.

* * *

En este documento estudiamos, en forma general, modelos de probabilidad de una variable aleatoria (v. a.) X , con funciones de distribución de uno o dos parámetros, con el enfoque mencionado más arriba. En particular, nos ocupamos de cómo ir poniendo al día sistemática y rigurosamente, siguiendo el paradigma bayesiano, el conocimiento incompleto acerca de los parámetros de una distribución dada que un investigador posee en cada momento, a partir de su información inicial y de la que va obteniendo a través de sucesivas muestras de la variable de interés, X . El objetivo final es una distribución de X coherente con (i) el modelo de probabilidad que gobierna X , que suponemos conocido, y (ii) la información de sus parámetros disponible en cada momento, obtenida a través de sucesivas muestras independientes de X .

Con todo ello, tendremos ocasión de ver «en funcionamiento y concretamente» las características básicas del proceso anticipadas más arriba.

Las secciones siguientes entran en detalle en los temas mencionados: la sección 2 es un breve repaso de unos cuantos conceptos generales útiles para entrar en materia a partir de la sección 3, que contiene los desarrollos teóricos fundamentales y la prueba de algunos teoremas de validez general. A modo de ilustración, las secciones 4 y 5 particularizan los resultados de la sección anterior a casos específicos de algunos de los modelos probabilísticos más utilizados. La esencia del documento, por consiguiente, reside en el contenido de la sección 3.

En particular, en las secciones 4 y 5 nos referiremos a las distribuciones que detallamos a continuación. A este efecto, distinguiremos entre:

- a) Funciones de densidad (f. d.) $\varphi(X | \theta)$, $\varphi(X | \alpha, \beta)$, donde X es una variable aleatoria continua de parámetros respectivos a θ y (α, β) , en general, parcialmente conocidos.
- b) Funciones de probabilidad (f. p.) $P(X | \theta)$ y $P(X | \alpha, \beta)$, donde X es una variable aleatoria discreta de parámetros respectivos a θ y (α, β) , en general, parcialmente conocidos.

En particular, estudiaremos los modelos siguientes:

- a) $\text{Ber}(X | p)$, dist. de Bernoulli, X v. a. discreta, $x = 0$ o $x = 1$; $p \in [0, 1]$
- b) $\text{Po}(X | \mu)$, dist. de Poisson, X v. a. discreta, $x \in \mathbb{N}$; $\mu \in \mathbb{R}^+$
- c) $\text{Exp}(X | \theta)$, dist. exponencial, X v. a. continua, $x \in \mathbb{R} \geq 0$; $\theta \in \mathbb{R}^+$
- d) $N(X | \mu, \sigma)$, dist. normal, X v. a. continua, $x \in \mathbb{R}$; $\mu \in \mathbb{R}$; $\sigma \in \mathbb{R}^+$

2. Conceptos generales: estadísticos y estimadores

En el resto del documento haremos referencia a algunos conceptos de carácter general que es conveniente resumir al principio; éste es el propósito de los párrafos siguientes en esta sección.

2.1. Algunos conceptos de interés

a. Muestra aleatoria genérica simple

Sean X_1, X_2, \dots, X_n variables aleatorias, independientes e idénticamente distribuidas (i. i. d.), de función de densidad φ o probabilidad P , entonces diremos que $\mathbf{X} = (X_1, X_2, \dots, X_n)$ constituye una **muestra aleatoria genérica simple** de tamaño n , generada por φ o P .

b. Muestra aleatoria simple observada

Llamaremos «muestra aleatoria simple» (m. a. s.) a cada una de las muestras $\mathbf{x} = (x_1, x_2, \dots, x_n)$ que constituyen una observación de la muestra genérica $\mathbf{X} = (X_1, X_2, \dots, X_n)$, en un experimento aleatorio específico.

c. Función de verosimilitud $L(\mathbf{X} | \theta)$ y $L(\mathbf{x} | \theta)$

Dada una m. a. s. genérica $\mathbf{X} = (X_1, X_2, \dots, X_n)$ de una población de función de densidad $\varphi(\mathbf{X} | \theta)$, llamaremos a la función «de densidad conjunta», que gracias a la independencia de los resultados viene definida por el siguiente producto:

$$\varphi(X_1 | \theta) \varphi(X_2 | \theta) \dots \varphi(X_n | \theta) = \varphi(\mathbf{X} | \theta),$$

función de verosimilitud de X_1, X_2, \dots, X_n dada θ , y la expresaremos mediante estas notaciones:

$$L(\mathbf{X} | \theta) = L[(X_1, X_2, \dots, X_n) | \theta] = \varphi(X_1 | \theta) \varphi(X_2 | \theta) \dots \varphi(X_n | \theta) = \varphi(\mathbf{X} | \theta)$$

Análogamente, si la función de densidad tiene dos parámetros α, β , la función de verosimilitud vendrá dada por:

$$L(\mathbf{X} | \alpha, \beta) = L[(X_1, X_2, \dots, X_n) | \alpha, \beta] = \varphi(X_1 | \alpha, \beta) \varphi(X_2 | \alpha, \beta) \dots \varphi(X_n | \alpha, \beta)$$

Cuando la variable aleatoria X es discreta de función probabilidad $P(\mathbf{X} | \theta)$ o $P(\mathbf{X} | \alpha, \beta)$, tendríamos, respectivamente:

$$L(\mathbf{X} | \theta) = L[(X_1, X_2, \dots, X_n) | \theta] = P(X_1 | \theta) P(X_2 | \theta) \dots P(X_n | \theta)$$

$$L(\mathbf{X} | \alpha, \beta) = L[(X_1, X_2, \dots, X_n) | \alpha, \beta] = P(X_1 | \alpha, \beta) P(X_2 | \alpha, \beta) \dots P(X_n | \alpha, \beta)$$

Obviamente, la función de verosimilitud de una muestra observada concreta respecto a θ o a (α, β) se obtiene sustituyendo en las expresiones anteriores $\mathbf{X} = (X_1, X_2, \dots, X_n)$ por $\mathbf{x} = (x_1, x_2, \dots, x_n)$.

Veremos que la función de verosimilitud juega un papel decisivo en el tratamiento del conocimiento relevante sobre los parámetros.

Observación. No debe confundirse $L(X_1, X_2, \dots, X_n | \theta)$, verosimilitud de la muestra (X_1, X_2, \dots, X_n) dada θ , con $L(\theta | X_1, X_2, \dots, X_n)$, verosimilitud de θ dada la muestra (X_1, X_2, \dots, X_n) .

d. Estadísticos

La función de densidad $\varphi(X | \theta_1, \theta_2, \dots, \theta_k)$ o de probabilidad $P(X | \theta_1, \theta_2, \dots, \theta_k)$ de una población raramente es totalmente conocida por el observador, porque su conocimiento de los parámetros $\theta_1, \theta_2, \dots, \theta_k$ es incompleto, por lo que generalmente está interesado en reducir su incertidumbre sobre dichos parámetros. Sus objetivos suelen ser:

i) Tener un conocimiento más completo de alguno de los parámetros $\theta_1, \theta_2, \dots, \theta_k$.

Téngase presente que estos parámetros suelen representar elementos muy importantes de la población representada por la variable X , como su media, proporción, varianza, etc. Para conocer mejor los parámetros $\theta_1, \theta_2, \dots, \theta_k$, el observador realiza un experimento aleatorio, obteniendo una m. a. s. numérica (x_1, x_2, \dots, x_n) . La condensación de la información obtenida se facilita reduciendo los datos, es decir, sustituyendo la muestra observada por **unos pocos números** calculados a partir de **una o varias funciones** aplicadas a la muestra (x_1, x_2, \dots, x_n) . Llamaremos «estadísticos» a estas funciones.

ii) Obtener mejores previsiones sobre los resultados de la variable X .

El objetivo (i) lo cubre la **inferencia estadística** y el (ii), el estudio de las **distribuciones predictivas**, que definiremos y estudiaremos más adelante.

Estadístico. Definición: llamaremos «estadístico» a toda función real:

$$d: \mathbf{X} = (X_1, X_2, \dots, X_n) \rightarrow \mathbb{R}$$

$$\mathbf{x} = (x_1, x_2, \dots, x_n) \rightarrow d(\mathbf{x}) = d(x_1, x_2, \dots, x_n),$$

donde X_1, X_2, \dots, X_n son «n» variables aleatorias independientes con i. i. d., y (x_1, x_2, \dots, x_n) es una muestra observada concreta.

e. Estadísticos conjuntamente suficientes

De un conjunto k de estadísticos

$$\{d_1(\mathbf{x}), d_2(\mathbf{x}), \dots, d_k(\mathbf{x})\}, \text{ donde } \mathbf{x} = (x_1, x_2, \dots, x_n),$$

diremos que es «**conjuntamente suficiente**» relativo a una muestra si contiene toda la información relevante a efectos de **inferencia y predicción**.

Teorema 1 (criterio de Fisher-Neyman)

Diremos que los k estadísticos $\{d_1(\mathbf{x}), d_2(\mathbf{x}), \dots, d_k(\mathbf{x})\}$ son conjuntamente suficientes para los parámetros θ o (α, β) de las funciones de probabilidad o densidad, respectivamente:

$$P(X | \theta), P(X | \alpha, \beta), \varphi(X | \theta) \text{ y } \varphi(X | \alpha, \beta),$$

si su función de verosimilitud puede ser expresada de esta forma:

$$L(\mathbf{x} | \theta) = h(\mathbf{x})g\{d_1(\mathbf{x}), d_2(\mathbf{x}), \dots, d_k(\mathbf{x}); \theta\}, \text{ para el caso de un solo parámetro, y}$$

$$L(\mathbf{x} | \alpha, \beta) = h(\mathbf{x})g\{d_1(\mathbf{x}), d_2(\mathbf{x}), \dots, d_k(\mathbf{x}); (\alpha, \beta)\}, \text{ para el caso de dos parámetros,}$$

donde h y g cumplen lo siguiente:

- a) son funciones no negativas;
- b) h sólo depende de \mathbf{x} , no de los parámetros, y

- c) g depende de los parámetros y también de \mathbf{x} , pero de \mathbf{x} sólo a través de los estadísticos $d_1(\mathbf{x}), d_2(\mathbf{x}), \dots, d_k(\mathbf{x})$.

Observación. Los estadísticos suficientes juegan un papel muy importante en la estadística. Ello es debido a que, como hemos dicho, a pesar de reducir los datos, «conservan toda la información relevante de la muestra a efectos de inferencia y predicción».

f. Estimador de un parámetro. Definición

Un estimador del parámetro θ de una población concreta es un estadístico «d» que, para toda m. a. s. $\mathbf{x} = (x_1, x_2, \dots, x_n)$ extraída de dicha población, da un valor aproximado θ' de dicho parámetro θ . Es decir, $d(\mathbf{x}) = \theta' \cong \theta$.

Dada una muestra m. a. s. específica $\mathbf{x} = (x_1, x_2, \dots, x_n)$, diremos que $d(\mathbf{x}) = \theta'$ es una estimación de θ .

Así, un estimador es el estadístico que, una vez obtenida una muestra, permite obtener una estimación θ' del parámetro θ .

g. Tipos de estimadores y propiedades

Definición 1

d es un estimador *insesgado* de θ si $E[d(\mathbf{x})] = \theta$.

Definición 2

Un estimador $d(\mathbf{x})$ es *consistente* si, para $n \rightarrow \infty$, $\lim d(\mathbf{x}) = \theta$.

Definición 3

El estimador $d_1(\mathbf{x})$ es *más eficiente* que $d_2(\mathbf{x})$ si $\text{Var}(d_1(\mathbf{x})) < \text{Var}(d_2(\mathbf{x}))$.

Definición 4

El estimador *más eficiente* es un estimador insesgado y de varianza mínima.

h. Algunos estadísticos suficientes importantes

Teorema 2

El par $(n, \sum x_i = r)$ define un estadístico conjuntamente suficiente relativo a una m. a. s., de tamaño n , generada por una variable aleatoria X de Bernoulli $\varphi_B(X | p) = p^X (1-p)^{1-X}$.

En efecto, la función de verosimilitud relativa a la muestra $\mathbf{x}(x_1, x_2, \dots, x_n)$ es

$$\begin{aligned} L(\mathbf{x} | p) &= \varphi_B(x_1 | p) \varphi_B(x_2 | p) \dots \varphi_B(x_n | p) = p^{x_1} (1-p)^{n-x_1} \cdot p^{x_2} (1-p)^{n-x_2} \dots p^{x_n} (1-p)^{n-x_n} = \\ &= p^r (1-p)^{n-r}, \end{aligned}$$

siendo

$$r = \sum x_i = d_1(\mathbf{x}), n = d_2(\mathbf{x}), h(\mathbf{x}) = 1.$$

Entonces,

$$L(\mathbf{x} | p) = p^r (1-p)^{n-r} = p^{d_1(\mathbf{x})} (1-p)^{d_2(\mathbf{x}) - d_1(\mathbf{x})},$$

y podemos escribir:

$$L(\mathbf{x} | p) = h(\mathbf{x})g(d_1(\mathbf{x}), d_2(\mathbf{x}), p),$$

$$\text{siendo } h(\mathbf{x}) = 1 \text{ y } g(d_1(\mathbf{x}), d_2(\mathbf{x}), p) = p^{d_1(\mathbf{x})} (1-p)^{d_2(\mathbf{x})-d_1(\mathbf{x})}.$$

Teorema 3

El par $(n, \sum x_i)$ define un estadístico conjuntamente suficiente relativo a una m. a. s. de tamaño n generada por la v. a. X de Poisson $Po(X | \mu) = e^{-\mu} \mu^x / x!$ ($x \in N$).

En efecto, la verosimilitud de la μ relativa a la muestra observada $\mathbf{x} = (x_1, x_2, \dots, x_n)$ es

$$L(\mathbf{x} | \mu) = \varphi_{Po}(\mathbf{x} | \mu) = \varphi_{Po}(x_1, x_2, \dots, x_n | \mu) = e^{-\mu n} \lambda^{\sum_{i=1}^n x_i} / (x_1! x_2! \dots x_n!) = h(\mathbf{x})g(d_1(\mathbf{x}), d_2(\mathbf{x}), \mu),$$

donde:

$$d_1(\mathbf{x}) = \sum x_i, d_2(\mathbf{x}) = n, h(\mathbf{x}) = 1/(x_1! x_2! \dots x_n!) \text{ y } g(d_1(\mathbf{x}), d_2(\mathbf{x}), \mu) = e^{-\mu n} \mu^{\sum x_i}$$

Teorema 4

El par $(n, \sum x_i)$ define un estadístico conjuntamente suficiente relativo a una m. a. s. (x_1, x_2, \dots, x_n) generada por la variable aleatoria X de densidad exponencial $\varphi_{Ex}(X | \mu) = (1/\mu)\exp(-X/\mu)$, siendo $\mu = E(X)$.

Ya que la verosimilitud relativa a la muestra observada $\mathbf{x} = (x_1, x_2, \dots, x_n)$ es

$$L(x_1, x_2, \dots, x_n | \theta) = \varphi_{Ex}(x_1, x_2, \dots, x_n | \theta) = k(\theta)^n \exp[-\theta (\sum x_i)],$$

si hacemos

$$h(\mathbf{x}) = 1, n = d_2(\mathbf{x}), \sum x_i = d_1(\mathbf{x}) \text{ y } g(d_1(\mathbf{x}), d_2(\mathbf{x}), \theta) = \theta^{d_2(\mathbf{x})} \exp[-\theta d_1(\mathbf{x})],$$

entonces queda:

$$\varphi_{Ex}(x_1, x_2, \dots, x_n | \mu) = h(\mathbf{x})g[d_1(\mathbf{x}), d_2(\mathbf{x}), \mu],$$

lo que prueba el teorema.

Teorema 5

El par $\{n, \sum x_i\}$ define un estadístico conjuntamente suficiente relativo a la m. a. s. X_1, X_2, \dots, X_n generada por la variable aleatoria X de distribución normal $N(X | \mu, \sigma)$ (siendo σ totalmente conocida).

La verosimilitud de la muestra observada $\mathbf{x} = (x_1, x_2, \dots, x_n)$ es

$$\begin{aligned} \varphi_N(x_1, x_2, \dots, x_n | \mu) &= \frac{1}{(2\pi)^{n/2} \sigma^n} \exp\left(-\frac{1}{2} \frac{\sum (x_i - \mu)^2}{\sigma^2}\right) = \\ &= \frac{1}{(2\pi)^{n/2} \sigma^n} \exp\left(-\frac{\sum x_i^2}{2\sigma^2}\right) \exp\left(\frac{n(2\mu \frac{\sum x_i}{n} - \mu^2)}{2\sigma^2}\right). \end{aligned}$$

Si hacemos

$$d_1(\mathbf{x}) = \frac{\sum x_i}{n}, d_2(\mathbf{x}) = n, h(\mathbf{x}) = \exp\left(-\frac{\sum x_i^2}{2\sigma^2}\right), k = 1/(2\sigma^2),$$

obtenemos:

$$g(d_1(\mathbf{x}), d_2(\mathbf{x}), \mu) = \frac{1}{(2\pi)^{d_2(\mathbf{x})/2} \sigma^{d_2(\mathbf{x})}} \exp\left\{kd_2(\mathbf{x})(2\mu d_1(\mathbf{x}) - \mu^2)\right\}$$

Por tanto, podemos escribir:

$$\varphi_N(\mathbf{x} | \mu) = h(\mathbf{x})g(d_1(\mathbf{x}), d_2(\mathbf{x}), \mu) \text{ con } h(\mathbf{x}) = 1,$$

lo que prueba el teorema 5.

Teorema 6

La terna $(\sum_{i=1}^n x_i^2, \sum_{i=1}^n x_i, n)$, **define un estadístico conjuntamente suficiente relativo a una muestra aleatoria simple de tamaño n, generada por la variable aleatoria X de distribución normal** $N(X | \mu, \tau)$, **con** $\mu = \mu_0$ **totalmente conocida y** $\tau = 1/\sigma^2$, **desconocida.**

Para su demostración, nos basaremos, como en los casos anteriores, en el teorema 1. La verosimilitud de la muestra $\mathbf{x}(x_1, x_2, \dots, x_n)$ generada por X viene dada por:

$$\begin{aligned} L(x_1, x_2, \dots, x_n | \tau) &= \frac{1}{2\pi\sigma^n} \exp\left\{-\frac{1}{2}\left[\left(\frac{x_1 - \mu_0}{\sigma}\right)^2 + \left(\frac{x_2 - \mu_0}{\sigma}\right)^2 + \dots + \left(\frac{x_n - \mu_0}{\sigma}\right)^2\right]\right\} = \\ &= \frac{1}{2\pi} \tau^{\frac{n}{2}} \exp\left\{-\frac{\tau}{2}[(x_1 - \mu_0)^2 + (x_2 - \mu_0)^2 + \dots + (x_n - \mu_0)^2]\right\} = \\ &= \frac{1}{2\pi} \tau^{\frac{n}{2}} \exp\left\{-\frac{\tau}{2}\left[\sum_{i=1}^n x_i^2 - 2\mu_0 \sum_{i=1}^n x_i + n\mu_0^2\right]\right\} \end{aligned}$$

Y si hacemos

$$h(\mathbf{x}) = 1/2\pi, \quad d_1(\mathbf{x}) = \sum_{i=1}^n x_i^2, \quad d_2(\mathbf{x}) = \sum_{i=1}^n x_i, \quad d_3(\mathbf{x}) = n,$$

podemos escribir:

$$L(\mathbf{x} | \tau) = \frac{1}{2\pi} \tau^{\frac{d_3(\mathbf{x})}{2}} \exp\left[-\frac{\tau}{2}(d_1(\mathbf{x}) - 2\mu_0 d_2(\mathbf{x}) + \mu_0^2 d_3(\mathbf{x}))\right]$$

Tal como queríamos probar, $L(\mathbf{x}, | \tau) = h(\mathbf{x})g[(d_1(\mathbf{x}), d_2(\mathbf{x}), d_3(\mathbf{x}), \tau]$

Observación importante. Conociendo totalmente $\mu = \mu_0$, aconsejamos calcular y guardar el estadístico $(\sum_{i=1}^n x_i^2, \sum_{i=1}^n x_i, n)$, ya que:

- Es un estadístico conjuntamente suficiente, pues cumple la condición de Fisher-Neyman.
- Podremos sumarlo fácilmente al análogo estadístico de una próxima muestra, para obtener así fácilmente el estadístico suficiente correspondiente a la unión de las dos muestras.
- A partir de $(\sum_{i=1}^n x_i^2, \sum_{i=1}^n x_i, n)$ podemos, conocido μ_0 , calcular fácilmente el *también estadístico suficiente* $(\sum_{i=1}^n (x_i - \mu_0)^2, n)$.

Teorema 7

La terna $(\sum_{i=1}^n x_i^2, \sum_{i=1}^n x_i, n)$, define un estadístico conjuntamente suficiente relativo a la m. a. s. (x_1, x_2, \dots, x_n) generada por la variable aleatoria X de función distribución $N(X | \mu, \sigma)$ con μ y σ parcialmente conocidas).

Demostración análoga a la de los teoremas anteriores:

Observación. El estadístico conjuntamente suficiente $(\sum_{i=1}^n x_i^2, \sum_{i=1}^n x_i, n)$ permite:

- a) Sumarlo fácilmente al análogo estadístico de una próxima muestra, para obtener así fácilmente el estadístico suficiente correspondiente a la unión de las dos muestras.
- b) Calcular fácilmente el estadístico suficiente $[(s^2 = \Sigma(x_i - m)^2 / v, m = (\Sigma x_i) / n \text{ y } v = n - 1)]$.

Resumen

Los siguientes son estadísticos conjuntamente suficientes relativos a una muestra observada (x_1, x_2, \dots, x_n) de las distribuciones indicadas:

<i>Función de observación (v. a.)</i>	<i>Estadístico suficiente</i>
Ber($X p$)	$(\sum_{i=1}^n x_i, n)$
Po($X \mu$)	$(\sum_{i=1}^n x_i, n)$
Exp($X \mu$)	$(\sum_{i=1}^n x_i, n)$
$N(X \mu, \sigma)$ (σ conocida totalmente)	$(\sum_{i=1}^n x_i, n)$
$N(X \mu, \tau)$ (μ conocida totalmente)	$(\sum_{i=1}^n x_i^2, \sum_{i=1}^n x_i, n), (\sum_{i=1}^n (x_i - \mu)^2, n)$
$N(X \mu, \tau)$ (μ y σ desconocidas)	$(\sum_{i=1}^n x_i^2, \sum_{i=1}^n x_i, n), (s^2, m, v)$

3. Distribuciones de los parámetros de los modelos generales $P(X | \theta), P(X | \alpha, \beta), \varphi(X | \theta)$ y $\varphi(X | \alpha, \beta)$

3.1. Introducción

En esta sección nos ocupamos de los modelos generales $P(X | \theta), P(X | \alpha, \beta), \varphi(X | \theta)$ y $\varphi(X | \alpha, \beta)$. Ya que el conocimiento del investigador sobre los parámetros es generalmente incompleto, nos preguntamos:

- ¿Cómo representa el investigador su conocimiento sobre ellos?
- ¿Asigna a los parámetros algún tipo de distribución?
- ¿Cómo llegar al tipo de distribución que represente correctamente su conocimiento?

El investigador desea generalmente expresar su conocimiento sobre los parámetros θ y (α, β) , después de obtener una muestra $\mathbf{x} = (x_1, x_2, \dots, x_n)$. Para ello:

- a) Considera los parámetros como variables aleatorias continuas.
- b) Expresa su conocimiento sobre los parámetros, conseguido a partir de muestras obtenidas anteriormente, mediante las correspondientes funciones de densidad $\varphi(\theta)$ o $\varphi(\alpha, \beta)$.

Estudiaremos las cuatro situaciones (A1, A2, B1 y B2) en que puede encontrarse el investigador:

A. El modelo $[P(X | \theta)$ o $\varphi(X | \theta)]$ tiene un solo parámetro $\theta \in \Theta$:

A1. El investigador tiene, previo a la muestra $\mathbf{x} = (x_1, x_2, \dots, x_n)$, un conocimiento incompleto de θ , representado por la función de densidad $\varphi(\theta)$.

A2. El investigador no tiene ningún conocimiento relevante del parámetro θ previo a la muestra.

B. El modelo $[P(X | \alpha, \beta)$ o $\varphi(X | \alpha, \beta)]$ tiene dos parámetros (α, β) :

B1. El investigador tiene de (α, β) un conocimiento inicial, previo a la muestra $\mathbf{x} = (x_1, x_2, \dots, x_n)$, representado por la función de densidad $\varphi(\alpha, \beta)$.

B2. El investigador no tiene ningún conocimiento relevante de los parámetros (α, β) previo a la muestra.

3.2. Modelos de un solo parámetro

Caso A1. El investigador resume su conocimiento sobre θ , inicial o previo a la muestra, mediante la función de densidad $\varphi(\theta)$

Distinguiremos dos casos:

a) **X continua:** de las relaciones

$$\varphi(\mathbf{x} \text{ y } \theta) = \varphi(\mathbf{x} | \theta)\varphi(\theta) = \varphi(\theta | \mathbf{x})\varphi(\mathbf{x}) \text{ y } \varphi(\mathbf{x}) = \int_{\Theta} \varphi(\mathbf{x} | \theta)\varphi(\theta)d\theta,$$

obtenemos la función de densidad $\varphi(\theta | \mathbf{x})$ de θ después de la muestra:

$$\varphi(\theta | \mathbf{x}) = \frac{\varphi(\mathbf{x} | \theta)\varphi(\theta)}{\varphi(\mathbf{x})} = \frac{\varphi(\mathbf{x} | \theta)\varphi(\theta)}{\int_{\Theta} \varphi(\mathbf{x} | \theta)\varphi(\theta)d\theta} = \frac{L(\mathbf{x} | \theta)\varphi(\theta)}{\int_{\Theta} L(\mathbf{x} | \theta)\varphi(\theta)d\theta},$$

siendo $\mathbf{x} = (x_1, x_2, \dots, x_n)$ y $\theta \in \Theta$.

b) **X discreta**

Análogamente al caso de X continua, en el caso de X discreta, de las relaciones

$$\varphi(\mathbf{x} \text{ y } \theta) = P(\mathbf{x} | \theta)\varphi(\theta) = \varphi(\theta | \mathbf{x})P(\mathbf{x}) \text{ y } P(\mathbf{x}) = \int_{\Theta} P(\mathbf{x} | \theta)\varphi(\theta)d\theta,$$

obtenemos la función de densidad del parámetro θ dada la muestra

$$\varphi(\theta | \mathbf{x}) = \frac{P(\mathbf{x} | \theta)\varphi(\theta)}{P(\mathbf{x})} = \frac{P(\mathbf{x} | \theta)\varphi(\theta)}{\int_{\Theta} P(\mathbf{x} | \theta)\varphi(\theta)d\theta} = \frac{L(\mathbf{x} | \theta)\varphi(\theta)}{\int_{\Theta} L(\mathbf{x} | \theta)\varphi(\theta)d\theta},$$

siendo $\mathbf{x} = (x_1, x_2, \dots, x_n)$ y $\theta \in \Theta$.

Resumiendo, en cualquiera de los casos, a) y b), si el investigador ha sabido explicitar su conocimiento sobre θ , previo a la muestra, mediante la correspondiente función de densidad $\varphi(\theta)$, entonces podemos enunciar:

Teorema 1

La estandarización del producto de la función de verosimilitud $L(\mathbf{x} | \theta) = L(x_1, x_2, \dots, x_n | \theta)$ por la función de densidad $\varphi(\theta)$ de θ , previa a la muestra, proporciona la densidad $\varphi(\theta | \mathbf{x}) = \varphi(\theta | x_1, x_2, \dots, x_n)$, resumen del conocimiento sobre θ después de la muestra.

Observación. En los apartados siguientes se probará que el tipo de distribución $\varphi(\theta)$ viene obligado por el tipo de distribución de la variable X .

Caso A2. El investigador no tiene conocimiento relevante sobre θ previo a la muestra

$\mathbf{x} = (x_1, x_2, \dots, x_n)$

En este caso, su conocimiento sobre el parámetro después de la muestra queda resumido por:

$$\varphi(\theta | \mathbf{x}) = \frac{L(\mathbf{x} | \theta)\varphi(\theta)}{\int_{\Theta} L(\mathbf{x} | \theta)\varphi(\theta)d\theta}.$$

Sin embargo, dado que el investigador no tiene ningún conocimiento relevante sobre θ , **su distribución a priori $\varphi(\theta)$ no depende de θ** , y por consiguiente:

$$\int_{\Theta} L(\mathbf{x} | \theta)\varphi(\theta)d\theta = \varphi(\theta) \int_{\Theta} L(\mathbf{x} | \theta)d\theta,$$

de modo que la expresión anterior puede escribirse así:

$$\varphi(\theta | \mathbf{x}) = \frac{L(\mathbf{x} | \theta)\varphi(\theta)}{\int_{\Theta} L(\mathbf{x} | \theta)\varphi(\theta)d\theta} = \frac{L(\mathbf{x} | \theta)\varphi(\theta)}{\varphi(\theta) \int_{\Theta} L(\mathbf{x} | \theta)d\theta} = \frac{L(\mathbf{x} | \theta)}{\int_{\Theta} L(\mathbf{x} | \theta)d\theta}.$$

Podemos así enunciar el siguiente teorema.

Teorema 2

Después de la muestra $\mathbf{x} = (x_1, x_2, \dots, x_n)$, la distribución $\varphi(\theta | \mathbf{x})$ que resume todo el conocimiento del investigador sobre θ , sin información relevante previa a dicha muestra, viene determinada por la estandarización de la función de verosimilitud $L(\mathbf{x} | \theta)$:

$$\varphi(\theta | \mathbf{x}) = \frac{L(\mathbf{x} | \theta)}{\int_{\Theta} L(\mathbf{x} | \theta)d\theta}$$

3.3. Modelos de dos parámetros

Caso B1. El investigador resume su conocimiento relevante sobre (α, β) , antes de la muestra, mediante la función de densidad $\varphi(\alpha, \beta)$

Teorema 3

La estandarización del producto de la función de verosimilitud $L[\mathbf{x} | \alpha, \beta]$ por la función de densidad $\varphi(\alpha, \beta)$, que resume el conocimiento del investigador previo a la obtención de la muestra $\mathbf{x} = (x_1, x_2, \dots, x_n)$, proporciona la función de densidad $\varphi(\alpha, \beta | \mathbf{x})$, después de la muestra:

$$\varphi(\alpha, \beta | \mathbf{x}) = \frac{L(\mathbf{x} | \alpha, \beta)\varphi(\alpha, \beta)}{\iint_C L(\mathbf{x} | \alpha, \beta)\varphi(\alpha, \beta)d\alpha d\beta} \quad (1),$$

donde C es el espacio de definición de (α, β) ; es decir, $(\alpha, \beta) \in C$.

Observación. Para llegar a este resultado basta con sustituir, en el teorema 1, θ , $d\theta$ y Θ , respectivamente, por (α, β) , $d\alpha d\beta$ y C .

Caso B2. El investigador no tiene, previo a la muestra, ningún conocimiento relevante sobre (α, β)

Simplemente generalizando el teorema 2 al caso de los parámetros, podemos enunciar:

Teorema 4

Si antes de la muestra $\mathbf{x} = (x_1, x_2, \dots, x_n)$ el investigador no tiene ningún conocimiento relevante sobre la distribución conjunta $\varphi(\alpha, \beta)$, su conocimiento después de la muestra $\mathbf{x} = (x_1, x_2, \dots, x_n)$ vendrá determinado por la estandarización de la función de verosimilitud:

$$\varphi(\alpha, \beta | \mathbf{x}) = \frac{L(\mathbf{x} | \alpha, \beta)}{\iint_C L(\mathbf{x} | \alpha, \beta) d\alpha d\beta},$$

donde C es el conjunto de definición de (α, β) .

Demostración. Suponemos que, previo a la primera muestra, el investigador no tiene ningún tipo de conocimiento sobre los parámetros (α, β) y, por tanto, que para él $\varphi(\alpha, \beta)$ no depende de (α, β) . Ello nos permite resumir su conocimiento sobre (α, β) , después de la primera muestra $\mathbf{x}(x_1, x_2, \dots, x_n)$, mediante:

$$\varphi(\alpha, \beta | \mathbf{x}) = \frac{L(\mathbf{x} | \alpha, \beta) \varphi(\alpha, \beta)}{\iint_C L(\mathbf{x} | \alpha, \beta) \varphi(\alpha, \beta) d\alpha d\beta} = \frac{L(\mathbf{x} | \alpha, \beta) \varphi(\alpha, \beta)}{\varphi(\alpha, \beta) \iint_C L(\mathbf{x} | \alpha, \beta) d\alpha d\beta} = \frac{L(\mathbf{x} | \alpha, \beta)}{\iint_C L(\mathbf{x} | \alpha, \beta) d\alpha d\beta}$$

Observación. Como en el caso A2, hemos podido sacar $\varphi(\alpha, \beta)$ fuera de la integral porque, dado el desconocimiento del investigador, $\varphi(\alpha, \beta)$ no depende de (α, β) .

3.4. Dadas dos muestras genéricas i. i. d. $X_A(X_{1A}, X_{2A}, \dots, x_{na})$ y $X_B(X_{1B}, X_{2B}, \dots, x_{mb})$ de $\varphi(X | \theta) \theta \in \Theta$, sin conocimiento relevante de θ previo a ellas, hallar $\varphi(\theta | X_A \cup X_B)$

Suponemos que, previo a las muestras X_A y X_B , no tenemos conocimiento sobre el parámetro θ . Podemos determinar la distribución final pedida $\varphi(\theta | X_A \cup X_B)$ por dos procedimientos distintos; analizaremos los resultados posteriormente.

a) Hallar $\varphi(\theta | X)$ directamente a partir de la muestra total $X = X_A \cup X_B$:

Ya que no se tiene información sobre X_A y X_B son independientes podemos escribir directamente:

$$\varphi(\theta | X) = \frac{L(X | \theta)}{\int_{\Theta} L(X | \theta) d\theta} = \frac{L(X_A \cup X_B | \theta)}{\int_{\Theta} L(X_A \cup X_B | \theta) d\theta} = \frac{L(X_A | \theta) L(X_B | \theta)}{\int_{\Theta} L(X_A | \theta) L(X_B | \theta) d\theta}$$

b) Procediendo secuencialmente, hallar $\varphi(\theta | X = X_A \cup X_B)$, conocidos $\varphi(\theta | X_A) = \psi(\theta)$ y la muestra X_B :

Conocida $\varphi(\theta | X_A) = \psi(\theta)$, después de la muestra X_B , la densidad final del parámetro θ vendrá dada por:

$$\pi(\theta | X_B) = \frac{\varphi(X_B | \theta) \psi(\theta)}{\int_{\Theta} \varphi(X_B | \theta) \psi(\theta) d\theta} = \frac{\psi(\theta) \varphi(X_B | \theta)}{\int_{\Theta} \psi(\theta) \varphi(X_B | \theta) d\theta} = \frac{\varphi(\theta | X_A) \varphi(X_B | \theta)}{\int_{\Theta} \varphi(\theta | X_A) \varphi(X_B | \theta) d\theta} \quad (1),$$

ya que

$$\varphi(\mathbf{X}_B | \theta) = L(\mathbf{X}_B | \theta) \quad (2)$$

y debido a no tener información previa a la muestra \mathbf{X}_A :

$$\varphi(\theta | \mathbf{X}_A) = \frac{L(\mathbf{X}_A | \theta)}{\int_{\Theta} L(\mathbf{X}_A | \theta) d\theta} \quad (3)$$

Sustituyendo (2) y (3) en (1) y simplificando, obtenemos:

$$\pi(\theta | \mathbf{X}_B) = \frac{\varphi(\theta | \mathbf{X}_A) \varphi(\mathbf{X}_B | \theta)}{\int_{\Theta} \varphi(\theta | \mathbf{X}_A) \varphi(\mathbf{X}_B | \theta) d\theta} = \frac{L(\mathbf{X}_A | \theta) L(\mathbf{X}_B | \theta)}{\int_{\Theta} L(\mathbf{X}_A | \theta) L(\mathbf{X}_B | \theta) d\theta}$$

Este resultado es igual al obtenido en el apartado a).

Observación importante. Se obtienen resultados análogos en el caso de dos parámetros.

3.5. Consecuencias importantes

Los resultados de los apartados a) y b) en 3.4. nos indican:

Después de la obtención de la muestra \mathbf{X}_B , tenemos *dos caminos* para llegar a la distribución de probabilidad que resumirá nuestro conocimiento sobre los parámetros:

- 1) Disponemos de los resultados de las muestras \mathbf{X}_A y \mathbf{X}_B y formamos una única muestra $\mathbf{X}_A \cup \mathbf{X}_B$. Antes de esta muestra $\mathbf{X}_A \cup \mathbf{X}_B$, el investigador no tenía ningún conocimiento previo de θ .
- 2) Una vez obtenida \mathbf{X}_B , utilizamos el conocimiento $\varphi(\theta | \mathbf{X}_A) = \psi(\theta)$, obtenido a través de la primera muestra \mathbf{X}_A , mediante la estandarización de la función de verosimilitud $L(\mathbf{X}_A | \theta)$. Llegamos al conocimiento final sobre θ , mediante la estandarización del producto $\psi(\theta)L(\mathbf{X}_B | \theta)$ (**teorema 1**). Hemos visto que los dos caminos conducen a un mismo resultado.
- 3) El tipo de distribución relativa a los parámetros **viene determinado** por la estandarización de la función de verosimilitud de una muestra genérica \mathbf{X} (es decir, del tipo de distribución que se ha supuesto para \mathbf{X}).
- 4) Si el investigador piensa que para llegar al conocimiento deseado de los parámetros de \mathbf{X} tendrá que realizar varias muestras, puede resultarle práctico y útil proceder según el método siguiente:
 - a) Después de la primera muestra $x_1(x_{11}, x_{12}, \dots, x_{1m})$, calcular y guardar su estadístico suficiente ES1.
 - b) Después de la segunda muestra $x_2(x_{21}, x_{22}, \dots, x_{2k})$, calculado su estadístico suficiente ES2, la muestra de unión $x_1 \cup x_2$ tendrá por estadístico suficiente ES1 + ES2.
 - c) De forma análoga, el estadístico suficiente de $x_1 \cup x_2 \cup x_3 \dots$ será ES1 + ES2 + ES3...

Calculados los estadísticos suficientes después de la primera, segunda, tercera muestra, etc., podrá calcular en cada caso la función de densidad que resume su conocimiento sobre los parámetros.

4. Distribuciones *a posteriori* de los parámetros de algunas distribuciones concretas

En esta sección aplicamos los resultados de la anterior para hallar la distribución de los parámetros de distribuciones concretas de X después de una muestra.

Distinguiremos dos casos:

- Sin información previa a la muestra.
- Con información previa a la muestra.

4.a. Distribuciones finales de los parámetros de algunas distribuciones concretas sin información previa a la muestra

4.a.1. Distribución del parámetro p de $Ber(x | p)$

Recordamos:

$$Ber(X | p) = p^X (1 - p)^{1 - X}$$

$X = 0$ si ocurre F (fallo), $X = 1$ si ocurre E (éxito) y $p = P(X = 1)$

La estandarización de la verosimilitud de la muestra proporcionará la distribución de p que resume el conocimiento del observador obtenido exclusivamente a partir de la muestra.

Teniendo en cuenta que:

- la muestra concreta $\mathbf{x}(x_1, x_2, \dots, x_n)$ está formada por una única sucesión de unos y ceros, y
- $P(1) = p$ y $P(0) = 1 - p$,

la verosimilitud $L(\mathbf{x} | p)$ será:

$$L(\mathbf{x} | p) = p^{x_1} (1 - p)^{1 - x_1} \cdot p^{x_2} (1 - p)^{1 - x_2} \dots p^{x_n} (1 - p)^{1 - x_n} = p^{\sum x_i} (1 - p)^{n - \sum x_i}$$

La **estandarización de la función de verosimilitud** $L(p | \mathbf{x})$ nos hace ver que la densidad de p , después de la m. a. s. $\mathbf{x}(x_1, x_2, \dots, x_n)$ y sin conocimiento sobre p previo, vendrá dada por una distribución de tipo *beta*:

$$Beta(p | a, b) = \frac{1}{B(a, b)} p^{a-1} (1 - p)^{b-1} = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} p^{a-1} (1 - p)^{b-1}$$

donde:

$$a - 1 = \sum x_i = r, \quad b - 1 = n - \sum x_i = n - r \quad \text{y} \quad a + b = n + 2$$

$$a = r + 1, \quad b = n - r + 1$$

siendo n = tamaño muestra, r = número de unos y $n - r$ = número de ceros.

Si el observador no tiene información previa a la muestra, el *estadístico suficiente* ($n, r = \sum x_i$), correspondiente a la muestra, determina los dos parámetros « $a = 1 + r$ » y « $b = n - r + 1$ », y, por tanto, la distribución de $Beta(p | a, b)$ después de la muestra.

4.a.2. Distribución del parámetro μ de $Po(x | \mu)$

Dada una m. a. s. $X(x_1, x_2, \dots, x_n)$ de la población $Po(x | \mu) = \frac{\mu^x \exp(-\mu)}{x!}$, la verosimilitud $L(\mathbf{x} | \mu)$ vendrá dada por:

$$L(\mathbf{x} | \mu) = \frac{\mu^{x_1} \exp(-\mu)}{x_1!} \cdot \frac{\mu^{x_2} \exp(-\mu)}{x_2!} \dots \frac{\mu^{x_n} \exp(-\mu)}{x_n!} = \frac{\mu^{\sum x_i} \exp(-n\mu)}{x_1! x_2! \dots x_n!} = C \mu^{\sum x_i} \exp(-n\mu)$$

donde $C = 1/(x_1! x_2! \dots x_n!)$.

La expresión de $L(\mathbf{x} | \mu)$ nos muestra que su estandarización nos llevará a una densidad para μ de tipo *gamma*:

$$\Gamma(\mu | a, b) = \frac{b^a}{\Gamma(a)} \mu^{a-1} \exp(-b\mu)$$

siendo $a = 1 + \sum x_i$ y $b = n$.

Si el observador no tiene información previa a la muestra, el *estadístico suficiente* ($n, \sum x_i$), correspondiente a la muestra, determina los dos parámetros « $a = 1 + \sum x_i$ » y « $b = n$ », y, por tanto, la distribución $\Gamma(\mu | a, b)$ de μ después de la muestra.

4.a.3. Distribución del parámetro θ de $Exp(x | \theta)$

Teniendo presente que:

$$\varphi_{Exp}(x | \theta) = \theta \exp(-\theta x) \quad (0 \leq x < +\infty, E(x) = 1/\theta = \mu; V(x) = 1/\theta^2),$$

la función de verosimilitud $L(\mathbf{x} | \theta)$ relativa a la m. a. s. $\mathbf{x}(x_1, x_2, \dots, x_n)$ viene dada por

$$L(\mathbf{x} | \theta) = \theta \exp(-\theta x_1) \theta \exp(-\theta x_2) \times \dots \times \theta \exp(-\theta x_i) \times \dots \times \theta \exp(-\theta x_n) = (\theta)^n \exp[-\theta(\sum x_i)]$$

La expresión de $L(\mathbf{x} | \theta)$ nos muestra que su estandarización conducirá a una densidad de θ de tipo *gamma*:

$$\Gamma(\theta | a, b) = \frac{b^a}{\Gamma(a)} \theta^{a-1} \exp(-b\theta)$$

siendo $a = n + 1$ y $b = \sum x_i = r$.

Si el observador no tiene información previa a la muestra, el *estadístico suficiente* ($n, r = \sum x_i$), correspondiente a la muestra, determina los dos parámetros « $a = n + 1$ » y « $b = r$ », y, por tanto, la distribución $\Gamma(\theta | a, b)$ de μ después de la muestra.

4.a.4. Distribución de μ de $N(x | \mu, \sigma)$ con σ conocida

Dada una muestra $\mathbf{x}(x_1, x_2, \dots, x_n)$ de una distribución $N(X | \mu, \sigma)$ de media μ desconocida y desviación tipo σ conocida, pretendemos hallar la distribución que puede atribuirse a μ .

Teniendo presente que:

$$\varphi_N(x_i | \mu) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x_i - \mu)^2}{2\sigma^2}\right\}$$

dada la m. a. s. (x_1, x_2, \dots, x_n) , la función de verosimilitud de la muestra, por ser σ constante, viene dada por (véanse las observaciones del final de párrafo):

$$\begin{aligned} L(\mathbf{x} | \mu) &\sim \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x_i - \mu)^2}{2\sigma^2}\right\} \sim \exp\left\{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}\right\} \sim \exp\left\{-\frac{1}{2\sigma^2} \left[\sum_{i=1}^n x_i^2 - 2\mu \sum_{i=1}^n x_i + n\mu^2\right]\right\} \\ &\sim \exp\left\{-\frac{n}{2\sigma^2} \left[-2\mu \frac{\sum_{i=1}^n x_i}{n} + \mu^2\right]\right\} \sim \exp\left\{-\frac{n}{2\sigma^2} (\mu^2 - 2\mu m)\right\} \sim \\ &\sim \exp\left\{-\frac{n}{2\sigma^2} (\mu^2 - 2\mu m)\right\} * \exp\left\{-\frac{n}{2\sigma^2} m^2\right\} \\ &\sim \exp\left\{-\frac{n}{2\sigma^2} (\mu^2 - 2\mu m + m^2)\right\} * \exp\left\{-\frac{n}{2\sigma^2} (\mu - m)^2\right\} \end{aligned}$$

La expresión de $L(\mathbf{x} | \mu)$ nos muestra que su estandarización conducirá a una densidad de μ de tipo *normal* $N(\mu | m, \sigma_\mu)$ de media $m = \sum x_i / n$ y desviación tipo $\sigma_\mu = \sigma / \sqrt{n}$.

Dada σ , el *estadístico suficiente* $\{n, r = \sum x_i\}$ determina los dos parámetros m y σ_μ , siendo $N(\mu | m, \sigma_\mu)$ la distribución final buscada.

Observaciones. Téngase en cuenta que a) « \sim » ha sido considerado como símbolo de proporcionalidad, y que b) la introducción de un segundo factor en la tercera fila es correcta, ya que es constante, es decir, no depende del parámetro μ .

4.a.5. Distribución de τ de $N(x | \mu, \tau = 1/\sigma^2)$ con μ conocida

Ya que $N(x_i | \mu, \tau) = C\tau^{1/2} \exp\left\{-\frac{\tau}{2}(x_i - \mu)^2\right\}$, la función de verosimilitud vendrá dada por:

$$L(\mathbf{x} | \tau) = k \tau^{n/2} \exp\left(-\frac{\tau}{2} \sum_{i=1}^n (x_i - \mu)^2\right)$$

La expresión de $L(\mathbf{x} | \tau)$ nos muestra que su estandarización conduce a una densidad de τ de tipo *gamma*:

$$\Gamma(\tau | a, b) = \frac{b^a}{\Gamma(a)} \tau^{a-1} \exp(-b\tau), \text{ siendo } a - 1 = \frac{n}{2} \text{ y } b = \frac{\sum_{i=1}^n (x_i - \mu)^2}{2}$$

Observación importante. Ya que $\sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n x_i^2 - 2\mu \sum_{i=1}^n x_i + n\mu^2$, el estadístico muestral suficiente $(\sum_{i=1}^n x_i^2, \sum_{i=1}^n x_i, n)$ determina, dada μ , los dos parámetros a y b , y, por tanto, la distribución final $\Gamma(\tau | a, b)$.

4.a.6. Distribución de (μ, τ) de $N(x | \mu, \tau)$, ambos desconocidos

Teniendo presente que:

$$N(x_i | \mu, \tau) = \frac{\tau^{1/2}}{\sqrt{2\pi}} \exp\left\{-\frac{\tau}{2}(x_i - \mu)^2\right\},$$

con μ y $\tau = 1/\sigma^2$ desconocidos, la función de verosimilitud vendrá dada por:

$$\begin{aligned} L(\mathbf{x} | \mu, \tau) &= k_1 \tau^{\frac{n}{2}} \exp\left\{-\frac{\tau}{2} \sum_{i=1}^n (x_i - \mu)^2\right\} = k_1 \tau^{\frac{n}{2}} \exp\left\{-\frac{\tau}{2} \sum_{i=1}^n [(x_i - m) + (m - \mu)]^2\right\} = \\ &= k_1 \tau^{\frac{n}{2}} \exp\left\{-\frac{\tau}{2} \left[\sum_{i=1}^n (x_i - m)^2 + n(m - \mu)^2 + 2(m - \mu) \sum_{i=1}^n (x_i - m) \right]\right\} = \text{(ver (*))} \\ &= k \tau^{\frac{1}{2}} \exp\left\{-\frac{n\tau}{2} (\mu - m)^2\right\} \times \tau^{\frac{n-1}{2}} \exp\left\{-\frac{\tau}{2} \sum_{i=1}^n (x_i - m)^2\right\} = \\ &= k \tau^{1/2} \exp\left\{-\frac{n\tau}{2} (\mu - m)^2\right\} \times \tau^{a-1} \exp\{-b\tau\} \end{aligned}$$

La expresión de $L(\mathbf{x} | \mu, \tau)$ nos muestra que la función de densidad de (μ, τ) será de tipo *normal-gamma*:

$$N\Gamma(\mu, \tau | n, m, a, b) = N(\mu | m, n\tau) \Gamma(\tau | a, b),$$

donde:

$$n = \text{tamaño muestral}, m = \frac{\sum_{i=1}^n x_i}{n}, a = \frac{n+1}{2}, b = \frac{\sum_{i=1}^n (x_i - m)^2}{2}$$

Observaciones:

1) (*) Hemos tenido presente para pasar a la igualdad siguiente que $\sum_{i=1}^n (x_i - m) = 0$.

2) Ya que

$$m = \left(\sum_{i=1}^n x_i\right) / n, \quad a = (n+1)/2 \quad \text{y} \quad \sum_{i=1}^n (x_i - m)^2 = \sum_{i=1}^n x_i^2 - 2m \sum_{i=1}^n x_i + nm^2$$

el estadístico conjuntamente suficiente $(n, \sum x_i, \sum x_i^2)$ determina los parámetros m , a y b , y, por tanto, la distribución final es de tipo *normal-gamma*.

3) Algunas relaciones de interés son las siguientes:

$$M(\mu) = m, \quad V(\mu) = \frac{\sigma^2}{n} = \frac{1}{n\tau}, \quad M(\tau) = \frac{a}{b}, \quad V(\tau) = \frac{a}{b^2}$$

4.a.7. Marginales de la distribución normal-gamma

Pretendemos hallar las distribuciones marginales de la distribución normal-gamma:

$$N\Gamma(\mu, \tau | n, m, a, b) = N(\mu | m, n\tau) \Gamma(\tau | a, b) = C \tau^{\frac{n}{2}} \exp\left\{-\left(b + \frac{n}{2}(\mu - m)^2\right)\tau\right\}$$

i. Distribución marginal $\varphi(\tau) = \Gamma(\tau | a, b)$

$$\varphi(\tau) = \int_{\mu=-\infty}^{\infty} N\Gamma(\mu, \tau | n, m, a, b) d\mu = \Gamma(\tau | a, b) \times \int_{\mu=-\infty}^{\infty} N(\mu | m, n\tau) d\mu = \Gamma(\tau | a, b) \times 1 = \Gamma(\tau | a, b)$$

$$M(\tau) = a/b, \quad V(\tau) = a/b^2$$

ii. Distribución marginal $\varphi(\mu)$

Ya que en $N(\mu | m, n\tau)$ es τ desconocida, no podemos dar directamente la distribución de μ .

Para hallarla, recordamos que cuando se muestrea una $N(x | \mu, \sigma)$, con μ y σ desconocidas, la variable aleatoria t

$$t = \frac{\mu - m}{\frac{s_x}{\sqrt{n}}} = \frac{\mu - m}{s_\mu} \quad (1)$$

sigue una distribución de Student con $\nu = n - 1$ grados de libertad, siendo:

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - m)^2}{n - 1} \quad \text{y} \quad s_\mu = \frac{s_x}{\sqrt{n}}$$

A partir de la distribuciones de Student, nos queda determinada la distribución marginal de μ . Representaremos las distribuciones de t y μ mediante estas notaciones:

$$St(t | \nu) \quad \text{y} \quad St(\mu | \nu, m, s_\mu)$$

iii. Resumen. Las marginales de la distribución $N\Gamma(\mu, \tau | n, m, a, b)$ son:

$$1) \quad St(\mu | \nu, m, s_\mu) \text{ con } \nu = n - 1, \quad m = \sum_{i=1}^n x_i / n, \quad s_\mu = \frac{s_x}{\sqrt{n}}, \quad M(\mu) = m \quad \text{y} \quad V(\mu) = s_\mu^2$$

$$2) \quad \Gamma(\tau | a, b) \text{ con } a = \frac{n+1}{2}, \quad b = \frac{1}{2} \sum_{i=1}^n (x_i - m)^2, \quad M(\tau) = a/b \quad \text{y} \quad V(\tau) = a/b^2$$

4.a.8. Relaciones entre estadísticos y parámetros

Puede observarse de las relaciones entre los parámetros y los estadísticos de las distribuciones estudiadas que:

- Dados los estadísticos suficientes, podremos calcular los parámetros (Relaciones I).
- Dados los parámetros, podremos calcular los estadísticos suficientes (Relaciones II).
- Para ello, ver la siguiente tabla:

<u>Dist. de X</u>	<u>Dist. Paramétrica</u>	<u>Relaciones I</u>	<u>Relaciones II</u>
Ber(X p)	Beta(p a,b)	$a = 1 + \sum x_i; b = n + 1 - \sum x_i$	$n = a + b - 2; \sum x_i = a - 1$
Po(X μ)	$\Gamma(\mu a,b)$	$a = 1 + \sum x_i; b = n$	$n = b; \sum x_i = a - 1$
Exp(X θ)	$\Gamma(\theta a,b)$	$a = n + 1; b = \sum x_i$	$n = 1 - a; \sum x_i = b$
N(X μ, σ), σ conocida	N($\mu m, \sigma_\mu$)	$m = (\sum x_i)/n; \sigma_\mu = \sigma/\sqrt{n}$	$n = \sigma^2/\sqrt{n}; \sum x_i = mn$
N(X μ, τ) μ conocida y $\tau = 1/\sigma^2$ desconocida	$\Gamma(\tau a,b)$	$a = (n/2) + 1; b = \{\sum(x_i - \mu)^2\}/2$	$n = 2a - 2; \sum(x_i - \mu)^2 = 2b$
N(X μ, τ) μ y τ desconocidas	$N\Gamma(\mu, \tau m, n, a, b) = N(\mu m, n\tau) \Gamma(\tau a, b)$	$m = (\sum x_i)/n$ $a = (n + 1)/2$ $b = \{\sum(x_i - \mu)^2\}/2$	$n = 2a - 1$ $\sum x_i = mn$ $\sum x_i^2 = 2[b + m^2(n - 1)]$

Observaciones importantes:

- Relaciones I nos muestran que, dado el estadístico suficiente relativo a una muestra, fácilmente podemos encontrar los parámetros.
- Si tenemos la correcta función de densidad de los parámetros que representa nuestro conocimiento sobre ellos, podemos hallar fácilmente el estadístico conjuntamente suficiente equivalente a dicho conocimiento (ver Relaciones II).

4.b. Distribuciones finales de los parámetros de algunas distribuciones concretas con información previa a la muestra

La información previa a la muestra actual $\mathbf{x}(x_1, x_2, \dots, x_n)$ puede venir dada por:

- Una muestra $\mathbf{x}_0(x_{10}, x_{20}, \dots, x_{n0})$ o su estadístico conjuntamente suficiente.
- Los parámetros $\theta_0, (\alpha_0, \beta_0)$ relativos a la distribución previa.

En el caso 1) podemos hallar los parámetros de la distribución final a partir de la suma de los estadísticos conjuntamente suficientes de las muestras \mathbf{x} y \mathbf{x}_0 .

En el caso 2), a partir de los parámetros que definen la distribución previa, podremos calcular el estadístico conjuntamente suficiente relativo a una muestra \mathbf{x}_0 equivalente al conocimiento dado por los parámetros (véase la tabla del apartado 4.a.8). Este estadístico conjuntamente suficiente sumado al estadístico conjuntamente suficiente relativo a la muestra \mathbf{x} dará el estadístico

conjuntamente suficiente, relativo a la muestra de unión de x y x_0 , que nos permitirá el cálculo de los parámetros buscados.

Distribución de los parámetros de una distribución (a partir del estadístico conjuntamente suficiente que resume el conocimiento del investigador)

Dist. de X	Estadístico suficiente	Dists. Paramétricas
Bin(X p)	(r = $\sum x_i$, n)	B(p a,b) a = r + 1, b = n - r + 1
Po(X μ)	(r = $\sum x_i$, n)	$\Gamma(\mu a,b)$ a = r + 1, b = n
Exp(X θ)	(r = $\sum x_i$, n)	$\Gamma(\theta a,b)$ a = n + 1, b = r
N(X μ, σ) σ conocida	(r = $\sum x_i$, n)	N($\mu m, \sigma_\mu$) m = ($\sum x_i$)/n , $\sigma_\mu = \sigma / \sqrt{n}$
N(X μ, σ) μ conocida	(n, $\sum x_i, \sum x_i^2$) o (n, $\sum_{i=1}^n (x_i - \mu)^2$)	$\Gamma(\tau a,b)$ { $\tau = 1/\sigma^2$ } a = (n/2) + 1 , b = $\sum (x - \mu)^2 / 2$
N(X μ, σ) μ y σ desconocidas ($\tau = 1/\sigma^2$)	(n, $\sum x_i, \sum x_i^2$)	N($\mu, \tau n, m, a, b$) = N($\mu m, n\tau$) $\Gamma(\tau a,b)$ Marginales St($\mu v, m, s_\mu$) y $\Gamma(\tau a,b)$ Siendo : $v = n - 1, m = \frac{\sum x_i}{n}, s_x^2 = \frac{\sum (x_i - m)^2}{n - 1}, s_\mu = \frac{s_x}{\sqrt{n}}, a = \frac{n + 1}{2}$ y $b = \frac{\sum (x_i - m)^2}{2}$

Observación importante. La tabla nos indica que, resumido nuestro conocimiento sobre una determinada población mediante una m. a. s. (x_1, x_2, \dots, x_n) o su estadístico suficiente, podemos determinar la distribución de sus parámetros.

* * *

Nótese que, como adelantábamos en general (sección 3), la acumulación de conocimiento sobre los parámetros se produce a través de sucesivas distribuciones de probabilidad que pertenecen a familias determinadas por las correspondientes distribuciones supuestas para X y que van variando según las muestras que se van acumulando.

5. Distribuciones predictivas de X distribuida según distintos modelos probabilísticos

En los procesos aleatorios de la variable X que nos ocupan, hemos supuesto que su función de densidad o de probabilidad es determinada por uno o varios parámetros, algunos de los cuales no son totalmente conocidos por el observador. Sin embargo, el observador puede asignar al parámetro o parámetros, gracias a la información incompleta que tiene sobre ellos, la correspondiente distribución. En esta sección, obtenemos, apoyándonos en dichas distribuciones, una distribución de X independiente de aquellos parámetros no totalmente conocidos pero coherentes con el conocimiento disponible acerca de los mismos en un momento determinado. A tal distribución de X la llamaremos «absoluta, no condicionada a los parámetros o predictiva».

Tendremos en cuenta que:

- 1) Si la v. a. X es *discreta*, su distribución predictiva viene definida por una función de probabilidad.
- 2) Si la v. a. X es *continua*, su distribución predictiva viene definida por una función de densidad de probabilidad.
- 3) Las v. a. X de las distribuciones binomial y de Poisson son discretas.
- 4) Las v. a. X de las distribuciones exponencial y normal son continuas.
- 5) El tipo de distribución de los parámetros, como hemos visto en secciones anteriores, viene obligado por el tipo de distribución de la v. a. X.

5.1. Distribución de la X de la binomial $\text{Bin}(X \mid p, n)$, cuando p sigue la distribución $\text{Beta}(p \mid a, b)$

El conocimiento que tiene un investigador sobre el parámetro p, proporción de individuos de la población que tienen cierta propiedad A, ha sido resumido mediante la distribución beta $B(p \mid a, b)$.

Dado el estado de su conocimiento sobre p, nos preguntamos: ¿Cuál es la probabilidad $P(X \mid n)$ de que en una muestra de n individuos, escogidos al azar, X de ellos tengan la propiedad A? O, dicho de otra manera, ¿cuál es la probabilidad de que en una muestra de n pruebas independientes, sacada de una población de $\text{Ber}(X \mid p)$, se haya obtenido X veces el valor 1?

Estas preguntas quedan contestadas mediante el teorema y el corolario que siguen.

Teorema

Si el conocimiento actual sobre p, parámetro de la distribución de $\text{Ber}(X \mid p)$, viene resumido por la distribución beta $B(p \mid a, b)$, la probabilidad de que en n pruebas independientes X tome x veces el valor 1 es:

$$P(X = x \mid n) = \binom{n}{x} \frac{B\{a+x, b+n-x\}}{B(a, b)},$$

que también puede escribirse, en función de la distribución gamma, como:

$$P(x | n) = \binom{n}{x} \times \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \times \frac{\Gamma(a+x)\Gamma(b+n-x)}{\Gamma(a+b+n)}$$

Observación. Esta probabilidad $P(X | n)$ no está condicionada a p ; podemos, por eso, llamarla «probabilidad absoluta, incondicional o predictiva» respecto a p .

Demostración. La probabilidad de x , dados n y p , es:

$$P(x | p, n) = \binom{n}{x} p^x (1-p)^{n-x},$$

y de $P(x | n) = \int_{p=0}^{p=1} P(x|p,n)P(p)dp$ obtenemos:

$$\begin{aligned} P(x|n) &= \int_{p=0}^{p=1} \binom{n}{x} p^x (1-p)^{n-x} P(p)dp = \int_{p=0}^{p=1} \binom{n}{x} p^x (1-p)^{n-x} \frac{1}{B(a,b)} p^{a-1} (1-p)^{b-1} dp = \\ &= \binom{n}{x} \frac{B\{a+x, b+n-x\}}{B(a,b)} \times \left(\frac{1}{B(a+x, b+n-x)} \int_{p=0}^{p=1} p^{(a+x)-1} (1-p)^{(b+n-x)-1} dp \right) = \\ &= \binom{n}{x} \frac{B\{a+x, b+n-x\}}{B(a,b)} \times 1 = \binom{n}{x} \frac{B\{a+x, b+n-x\}}{B(a,b)} \end{aligned}$$

5.2. Distribución de la X de Poisson $Po(X | \mu)$ cuando el conocimiento de μ viene resumido por $\Gamma(\mu | a, b)$

Supongamos que el observador resume su conocimiento sobre la media μ , de un proceso de Poisson $Po(X | \mu)$, mediante la función de densidad $\Gamma(\mu | a, b)$, donde a y b son constantes conocidas.

En estas circunstancias, se pregunta acerca de la probabilidad de que en un experimento el resultado de dicho proceso de Poisson sea $X = x$.

La respuesta viene dada por el teorema que sigue:

Teorema

Si el conocimiento actual sobre μ es resumido por la distribución gamma $\Gamma(\mu | a, b)$, la probabilidad de que en un experimento de Poisson de distribución $Po(X | \mu)$ la variable aleatoria X tome el valor x viene dada por la función de probabilidad absoluta, no condicionada o predictiva.

$$P(X = x) = \frac{b^a \Gamma(a+x)}{x!(b+1)^{a+x} \Gamma(a)}$$

Demostración. En efecto, teniendo presente que:

$$P(x | \mu) = \frac{\mu^x \exp(-\mu)}{x!} \quad \text{y} \quad \varphi_{\Gamma}(\mu | a, b) = \frac{b^a \mu^{a-1} \exp(-b\mu)}{\Gamma(a)}$$

la probabilidad «no condicionada» de x es:

$$\begin{aligned} \text{Po}(x) &= \int_{\mu=0}^{\infty} \text{Po}(x | \mu) \varphi_{\Gamma}(\mu | a, b) d\mu = \int_{\mu=0}^{\infty} \frac{\mu^x \exp(-\mu)}{x!} \times \frac{b^a \mu^{a-1} \exp(-b\mu)}{\Gamma(a)} d\mu = \\ &= \frac{b^a}{x! \Gamma(a)} \int_{\mu=0}^{\infty} \mu^{x+a-1} \exp\{-(b+1)\mu\} d\mu \end{aligned}$$

Con el cambio de variable

$$(b+1)\mu = \mu_1, \quad d\mu = \frac{1}{b+1} d\mu_1$$

obtenemos:

$$\begin{aligned} \text{Po}(x) &= \frac{b^a}{x! \Gamma(a)} \times \frac{1}{(b+1)^{x+a}} \int_{\mu_1=0}^{\infty} \mu_1^{x+a-1} \exp\{-\mu_1\} d\mu_1 = \frac{b^a}{x! \Gamma(a)} \times \frac{1}{(b+1)^{x+a}} \Gamma(x+a) = \\ &= \frac{b^a}{x!(b+1)^{x+a}} \times \frac{\Gamma(x+a)}{\Gamma(a)} \end{aligned}$$

Tal como queríamos probar.

Corolario

Siendo la información de que dispone el investigador la obtenida por una muestra X_1, X_2, \dots, X_n de tamaño n , generada por la f. de o. X , entonces la distribución de probabilidad predictiva de $X = x$ viene dada por:

$$\text{Po}(x) = \frac{b^a \Gamma(a+x)}{x!(b+1)^{a+x} \Gamma(a)}, \quad \text{donde } a = n \text{ y } b = \sum X_i.$$

5.3. Densidad $\varphi(x)$ de la X de $\text{Exp}(X | \theta)$ cuando el conocimiento de θ viene resumido por $\Gamma(\theta | a, b)$

Las preguntas análogas a las hechas en 5.2., pero referidas ahora a procesos exponenciales cuya v. a. X es continua, enunciamos el teorema siguiente:

Teorema

Si el conocimiento relativo al parámetro θ de una distribución exponencial de densidad $\varphi_{\text{Exp}}(X | \theta) = \theta \exp(-\theta X)$ viene resumido por la distribución gamma $\Gamma(\theta | a, b)$, con a y b conocidas, entonces la funciones de densidad y de probabilidad acumulada predictivas de la v. a. X serán, respectivamente:

$$\varphi_{\text{Exp}}(x) = \frac{ab^a}{(b+x)^{a+1}} \quad \text{y} \quad F_{\text{Exp}}(x) = 1 - \left[\frac{b}{b+x} \right]^a$$

Demostración. Teniendo presente que:

$$\varphi_{\text{Exp}}(X | \theta) = \theta \exp(-\theta X),$$

la densidad predictiva o no condicionada de x es:

$$\begin{aligned} \varphi_{\text{Exp}}(x) &= \int_{\theta=0}^{\infty} \varphi_{\text{Exp}}(x | \theta) \varphi_{\Gamma}(\theta | a, b) d\theta = \int_{\theta=0}^{\infty} \theta \exp(-\theta x) \frac{b^a}{\Gamma(a)} \theta^{a-1} \exp(-b\theta) d\theta = \\ &= \frac{b^a}{\Gamma(a)} \int_{\theta=0}^{\infty} \theta^a \exp\{-\theta(x + b)\} d\theta \end{aligned}$$

Con el cambio de variable

$$\theta(b + x) = \theta_1, \quad \theta = \frac{\theta_1}{b + x},$$

obtenemos la función de densidad:

$$\varphi_{\text{Exp}}(x) = \frac{b^a}{(b + x)^{a+1} \Gamma(a)} \int_{\theta_1=0}^{\infty} \theta_1^a \exp\{-\theta_1\} d\theta_1 = \frac{b^a}{(b + x)^{a+1} \Gamma(a)} \times \Gamma(a + 1) = \frac{ab^a}{(b + x)^{a+1}}$$

que integrando entre 0 y x da la función de probabilidad acumulada no condicionada:

$$F_{\text{Exp}}(x) = \int_0^x \frac{ab^a}{(b + u)^{a+1}} du = 1 - \left[\frac{b}{b + x}\right]^a$$

donde $a = n$ y $b = \sum x_i$, como enunciábamos en el teorema.

5.4. Distribución de X de $N(X | \mu, \sigma)$

Distinguiremos tres casos:

- a) μ parcialmente conocida y σ totalmente conocida
- b) μ totalmente conocida y σ parcialmente conocida
- c) μ y σ parcialmente conocidas

Caso a. σ totalmente conocida

Suponiendo que el conocimiento del observador sobre la media μ del proceso normal $N(X | \mu, \sigma)$ viene dado por la distribución $N(\mu | \mu_1, \sigma_1)$ de parámetros μ_1 y σ_1 completamente conocidos, nos preguntamos cuál es la función de densidad de X , obtenida del proceso $N(X | \mu, \sigma)$.

Teorema

Si X sigue la distribución normal $N(X | \mu, \sigma^2)$ con σ totalmente conocida, y el conocimiento sobre μ viene descrito por la distribución $N(\mu | \mu_1, \sigma_1^2)$, donde μ_1 y σ_1 son constantes conocidas, la distribución predictiva de X es la normal $N[X | \mu_1, (\sigma_1^2 + \sigma^2)]$.

Demostración. Si σ es totalmente conocida y μ tiene la distribución $N(\mu | \mu_1, \sigma_1^2)$ de parámetros también totalmente conocidos, la densidad no condicionada o predictiva de x viene dada por $\varphi(X) =$.

Después de sustituir las respectivas densidades, agrupar convenientemente e integrar, se obtiene:

$$\varphi(X) = [2\pi (\sigma_1^2 + \sigma^2)]^{-(1/2)} \exp\left\{-\frac{(x - \mu)^2}{2(\sigma_1^2 + \sigma^2)}\right\} = \varphi_N(X | \mu_1, \sigma_1^2 + \sigma^2),$$

lo que prueba el enunciado.

Resumen

- 1) La distribución de X es $N(X | \mu, \sigma)$, σ totalmente conocida y μ parcialmente conocida.
- 2) El parámetro μ tiene asignada la distribución $N(\mu | \mu_1, \sigma_1^2)$.
- 3) La distribución predictiva de X es la normal $N(X | \mu_1, \sigma_1^2 + \sigma^2)$.
- 4) La media de la predictiva de X es $\mu_1 = E(X) = E(\mu)$.
- 5) La varianza de la predictiva de X es $\sigma_1^2 + \sigma^2 = \text{Var}(\mu) + \text{Var}(X)$.

Corolario

Si la única información de que dispone el investigador es la obtenida a partir de una muestra $(x_1, x_2, x_3, \dots, x_n)$ generada por la función de observación X de distribución $N(X | \mu, \sigma^2)$ (con μ desconocida y σ conocida), la distribución predictiva o incondicional de X es la normal $N\{X | m, \sigma^2[1 + (1/n)]\}$.

En efecto, supuesta una distribución inicial no informativa para μ , la información proporcionada por la muestra nos permite pasar a la *a posteriori* $N(\mu | \mu_1, \sigma_1^2)$ con $\mu_1 = m = (\sum x_i)/n$ y $\sigma_1^2 = \sigma^2/n$. Sustituyendo estos valores en el teorema anterior, obtenemos la distribución predictiva del corolario.

Observación importante. Obtenidas las distribuciones predictivas correspondientes al teorema y el corolario anteriores, se podrán hallar fácilmente intervalos de confianza para las correspondientes variables aleatorias X normales.

Caso b. μ totalmente conocida

Si el conocimiento del investigador sobre $\tau = 1/\sigma^2$, de la distribución normal $N(X | \mu, \tau)$, viene dado por la distribución $\Gamma(\tau | a, b)$, donde a y b son constantes conocidas, nos preguntamos cuál es la probabilidad de que $x \leq k$.

Teorema

Si la v. a. X se distribuye según $N(X | \mu, \tau)$, con μ conocida, y el conocimiento de τ viene descrito por la distribución $\Gamma(\tau | a, b)$, donde a y b son constantes conocidas y $\sigma_1^2 = b/a$, $v_1 = 2a$ y $Z = (X - \mu)/\sigma_1$, entonces X y Z tienen las siguientes distribuciones de Student:

$$\text{St}(X | \mu, \sigma_1, v_1) \quad \text{y} \quad \text{St}(Z | v_1)$$

Demostración. La función de densidad incondicional o predictiva se obtiene de:

$$\varphi_N(X) = \int_{\tau=0}^{\infty} \varphi_N(X | \mu, \tau) \varphi_{\Gamma}(\tau | a, b) d\tau.$$

Sustituyendo $\varphi_N(X | \mu, \tau)$ y $\varphi_{\Gamma}(\tau | a, b)$ por sus expresiones e integrando, después de algunas transformaciones se obtiene:

$$\varphi_N(X) = \varphi_{St}(X | \mu, \sigma_1, \nu_1),$$

donde:

$$\sigma_1^2 = b/a \quad \text{y} \quad \nu_1 = 2a,$$

y el cambio $Z = (X - \mu)/\sigma_1$ transforma $\varphi_{St}(X | \mu, \sigma_1, \nu_1)$ en $\varphi_{St}(Z | \nu_1)$.

Estas distribuciones permiten calcular $P(X \leq k) = P(Z \leq (k - \mu)/\sigma_1)$.

Caso c. μ y σ parcialmente conocidas

Si el conocimiento previo del observador sobre μ y $\tau = 1/\sigma^2$ en un proceso normal $N(X | \mu, \tau)$ viene definido por la función de densidad

$$N\Gamma(\mu, \tau | \mu_1, n_1, a, b) = N(\mu | \mu_1, n_1 \tau) \Gamma(\tau | a, b),$$

donde μ_1, n_1, a y b son constantes conocidas, nos preguntamos cuál es la probabilidad de que la v. a. X del proceso normal $N(X | \mu, \tau)$ cumpla $X \leq k$.

Teorema

Si X sigue la distribución $N(X | \mu, \tau)$, el conocimiento sobre (μ, τ) viene descrito por la distribución conjunta

$$N\Gamma(\mu, \tau) = N(\mu | \mu_1, n_1 \tau) \Gamma(\tau | a, b),$$

donde μ_1, n_1, a y b son constantes conocidas por el observador, y

$$\sigma_1^2 = \frac{b(n_1 + 1)}{a n_1}, \quad \nu_1 = 2a \quad \text{y} \quad Z = \frac{X - \mu_1}{\sigma_1},$$

entonces X y Z tienen las siguientes distribuciones:

$$St(X | \mu, \sigma_1, \nu_1) \quad \text{y} \quad St(Z | \nu_1).$$

Demostración. La función de densidad predictiva o incondicional la calculamos a partir de:

$$\varphi(X) = \int_{\mathbb{R}^+} \int_{\mathbb{R}} \varphi_N(X | \mu, \tau) \varphi_N(\mu | \mu_1, n_1 \tau) \varphi_{\Gamma}(\tau | a, b) d\mu d\tau.$$

Sustituyendo $\varphi_N(x | \mu, \tau)$, $\varphi_N(\mu | \mu_1, n_1 \tau)$ y $\varphi_{\Gamma}(\tau | a, b)$ por sus expresiones e integrando, se obtiene, después de algunas transformaciones:

$$\varphi(X) = \varphi_{St}(X | \mu_1, \sigma_1, \nu_1),$$

donde μ_1 es un dato, $\sigma_1^2 = \frac{b(1 + n_1)}{a n_1}$ y $\nu_1 = 2a$.

El cambio $Z = \frac{X - \mu_1}{\sigma_1}$ transforma $\varphi_{St}(X | \mu_1, \sigma_1, \nu_1)$ en $\varphi_{St}(Z | \nu_1)$.

Estas distribuciones permiten calcular

$$P(X \leq k) = P\left(Z \leq \frac{k - \mu_1}{\sigma_1}\right)$$

Corolario. Sea la muestra (x_1, x_2, \dots, x_n) , generada por la f. de o. X , la única información de que dispone el investigador de los parámetros (μ, σ) , dicha información se resume, como sabemos, mediante la distribución

$$N\Gamma(\mu, \sigma) = N(\mu | \mu_1, n_1 \tau) \Gamma(\tau | a, b),$$

con $\mu_1 = m = \Sigma x_i$, $a = (n - 1)/2$, $b = \Sigma(x_i - m)^2/2$, $n_1 = n$.

Entonces, las distribuciones predictivas o incondicionales de X y $Z = (X - \mu_1)/\sigma_1$ son $St(X | \mu_1, \sigma_1, \nu_1)$ y $St(Z | \nu_1)$, con $\mu_1 = m$, $s^2 = \Sigma(x_i - m)^2/\nu_1$, $\sigma_1^2 = (1 + (1/n))s^2$, $\nu_1 = n - 1$.

En efecto, sustituyendo los valores del corolario en el teorema anterior, obtenemos:

$$\nu_1 = 2a = n - 1, \mu = m$$

$$\sigma_1^2 = \frac{b(1 + n_1)}{an_1} = \left(1 + \frac{1}{n}\right) \frac{\Sigma(x_i - m)^2}{n - 1} = \left(1 + \frac{1}{n}\right) s^2$$

DISTS. INICIALES DE X	DISTS. DE LOS PARÁMETROS	DISTS. PREDICTIVAS DE X
Bin(x p,n)	Beta(p a,b) a = $\sum x_i + 1$, b = n - $\sum x_i + 1$ n = tamaño muestral	$P(x n) = \binom{n}{x} \frac{B(a+x, b+n-x)}{B(a, b)}$
Po(x μ)	$\Gamma(\mu a,b)$ a = 1 + $\sum x_i$, b = n	$P(x) = \frac{b^a \Gamma(a+x)}{x!(b+1)^{a+x} \Gamma(a)}$
Exp(x θ)	$\Gamma(\theta a,b)$ a = n + 1, b = $\sum x_i$	$\varphi(x) = \frac{ab^a}{(b+x)^{a+1}}, F(x) = 1 - \left(\frac{b}{b+x}\right)^a$
N(x μ, σ), σ conocida	N($\mu \mu_1, \sigma_1$) $\mu_1 = (\sum x_i / n) = m$, $\sigma_1 = \sigma^2 / n$	$N\left\{x m, \sigma^2 \left(1 + \frac{1}{n}\right)\right\}$
N(x μ, σ) μ conocida	$\Gamma(\tau a, b)$ a = n/2, b = $\sum (x_i - \mu)^2 / 2$	St(x μ, σ_1, ν_1) $\sigma_1^2 = b/a$, $\nu_1 = 2a$
N(x μ, σ) μ y σ desconocidas	$\mu_1 = m$, $s^2 = \sum (x_i - m)^2 / \nu_1$	S(x μ_1, σ_1, ν_1) $\mu_1 = m$, $\nu_1 = n - 1$, $\sigma_1^2 = \left(1 + \frac{1}{n}\right) s^2$

Referencias

Fischhoff, B., P. Slovic y S. Lichtenstein (1988), «Knowing what you want: measuring labile values», en D. Bell, H. Raiffa y A. Tversky (eds.): *Decision-making. Descriptive, normative and prescriptive interactions*, Cambridge University Press, Cambridge, capítulo 18.