

Departament d'Estadística i I.O., Universitat de València.
Facultat de Matemàtiques, 46100–Burjassot, València, España.
Tel. +34.96.364.3560 (directo), +34.96.386.4362 (secretaría)
Fax +34.96.364.3560 (directo), +34.96.386.4735 (secretaría)
E-mail: jose.m.bernardo@uv.es, Web: <http://www.uv.es/~bernardo/>

Compuesto el 13 de Enero de 2000

Métodos Estadísticos Contemporáneos en la Investigación Científica

JOSÉ M. BERNARDO
Universitat de València, Spain

Contemporary Statistical Methods in Scientific Research

A new, extremely powerful paradigm to analyse the results of scientific experimentation, which is firmly based on axiomatic foundations, is gradually substituting the traditional recipe-based techniques which have been dominant in statistics for most of this century. The new paradigm, usually referred to as the *Bayesian* methodology, uses a probability concept which closely matches that used in ordinary language, directly solves the more relevant scientific questions on data analysis, and may be applied to complex, richly structured problems, fairly inaccessible to traditional statistical methods.

Introducción. Los métodos estadísticos convencionales han sido durante generaciones una fuente constante de insatisfacción, tanto para los estudiantes universitarios como para los científicos en activo que, aunque muy interesados en un análisis adecuado de sus datos experimentales, no se sienten especialmente atraídos por una colección de recetas claramente *carentes* de una estructura lógica que las relacione; obviamente, científicos o ingenieros encuentran útiles formularios que les resuman los procedimientos de uso más frecuente pero, entrenados en un razonamiento lógico, exigen que se les ofrezca una justificación convincente. Convencionalmente, la formación en probabilidad y estadística da comienzo en bachillerato con ejemplos elementales de probabilidad, que ciertamente resultan razonables, pero que *parecen* limitarse a problemas relacionados con los juegos de azar. Naturalmente, los alumnos con aspiraciones a un trabajo científico, tanto en las ciencias de la naturaleza como en las ciencias sociales, se muestran interesados por aprender a analizar datos experimentales; sin embargo, suelen sentirse decepcionados cuando, ya en la universidad, se les empieza a hablar de inferencia estadística. En efecto, aunque muchos de los procedimientos que se les exponen parecen sensatos por separado, no se les pone de manifiesto (porque *no existe*) ningún *principio* básico que los estructure. Los cursos de probabilidad y estadística parecen llevar a una dicotomía perversa: la teoría de la probabilidad está bien fundamentada, pero sólo parece aplicable a problemas de interés muy restringido; los problemas de inferencia estadística son científica y socialmente importantes, pero su ‘teoría’ se reduce a una colección de procedimientos, posiblemente intuitivos, pero que no sólo carecen de una justificación *conjunta*, sino que además son frecuentemente *incompatibles* entre sí. Lo que rara vez se les explica a los estudiantes es que *existe* un paradigma alternativo, la *metodología bayesiana*, que únicamente requiere las *matemáticas* de una teoría general de la probabilidad perfectamente justificada, y el *concepto* de probabilidad

que corresponde al uso convencional de esa palabra en el lenguaje cotidiano. Las consecuencias prácticas de la adopción de este nuevo paradigma son de muy largo alcance: por una parte, los métodos bayesianos reducen los problemas de inferencia estadística a problemas de la teoría de la probabilidad, lo que hace esencialmente innecesaria la introducción de nuevos conceptos; por otra parte, la nueva metodología proporciona una *justificación lógica* a muchos de los procedimientos estadísticos convencionales como *aproximaciones* adecuadas (especialmente en el caso de grandes muestras), haciendo explícitas las *condiciones* que resultan necesarias para garantizar su validez. En el limitado espacio de que dispongo, intentaré sintetizar, sin ninguna pretensión de profundidad matemática, los elementos básicos de los métodos estadísticos bayesianos.

Los problemas planteados. Frecuentemente, los resultados experimentales están constituidos por un conjunto $D = \{x_1, \dots, x_n\}$ de $n \geq 1$ observaciones *homogéneas* x_i (la palabra técnica es *intercambiables*), posiblemente multidimensionales; el científico se interesa entonces por lo que puede *inferirse*, en base a tales datos, sobre la *naturaleza* del proceso estudiado, o sobre el comportamiento de observaciones *futuras* del mismo proceso. La respuesta a tales preguntas requiere la especificación de un *modelo experimental* que describa la *relación* existente entre los datos disponibles, D , y la *naturaleza* del proceso, frecuentemente descrita por una cierta magnitud ω (estado de la naturaleza) sobre la que se desea obtener información. Como ejemplos, muy elementales, consideraremos (i) el resultado positivo $D = \{+\}$, obtenido al efectuar a una persona elegida al azar un test de tuberculina diseñado para la detección precoz de la tuberculosis; (ii) el conjunto $D = \{9.78, 9.84, 9.97, 10.02, 9.75, 9.68, 9.84, 9.78, 9.98, 9.73\}$, de 10 medidas independientes, en m/seg^2 , de la constante gravitatoria g , realizadas con un péndulo en unas prácticas de la asignatura de mecánica; (iii) los resultados $D = \{0, 0, \dots, 0\}$, de los tests de inmuno-deficiencia realizados a 200 estudiantes, que han sido felizmente negativos para todos ellos; y (iv) las presiones máximas $D = \{59.5, 59.8, 63.5, 57.1, 65.0\}$, en kg/cm^2 , que han resistido 5 piezas extraídas al azar entre las que forman parte de un lote industrial, cuyos elementos sólo son aceptables si resisten al menos 50 kg/cm^2 . En el primer caso queremos determinar la proporción de personas con tuberculosis que puede esperarse entre las que han dado positivo en un test de tuberculina; en el segundo ejemplo, interesan las conclusiones que puedan inferirse de los datos obtenidos en el laboratorio de prácticas sobre el verdadero valor de la constante gravitatoria g ; en el tercero, pretendemos extraer conclusiones sobre la proporción de seropositivos entre los estudiantes; en el último ejemplo, queremos sacar conclusiones sobre la proporción de piezas en el lote estudiado que, previsiblemente, no cumplirán la especificación requerida.

El concepto de probabilidad. El elemento característico del paradigma bayesiano es su concepto de *probabilidad como medida de incertidumbre*, precisamente en el mismo sentido en el que la palabra ‘probabilidad’ se utiliza en el lenguaje cotidiano. Así, cuando en el primero de los ejemplos anteriores, afirmemos que la *probabilidad* de que una persona a quien el test de tuberculina haya dado positivo, esté efectivamente enferma de tuberculosis es 0.165, estaremos afirmando que puede esperarse que tan sólo el 16.5% de las personas que han dado resultado positivo ante el test de tuberculina (y que como consecuencia serán exploradas con detalle) resultarán estar realmente enfermas de tuberculosis. En el segundo ejemplo, cuando afirmemos que *en base a los resultados experimentales obtenidos* la probabilidad de que el valor de constante gravitatoria g esté entre 9.75 y 9.92 es de 0.95, estaremos cuantificando la incertidumbre residual, después de analizar los datos, sobre el *verdadero* valor de g , que estamos suponiendo desconocido; naturalmente, puesto que se trata de una constante física, el suceso de que g se sitúe entre 9.75 y 9.92 es verdadero o es falso (de hecho, es verdadero); sin

embargo, lo *único* que los datos experimentales permiten afirmar sobre ese suceso es que es *probablemente* cierto, con el valor 0.95 como *medida* de tal probabilidad. De forma análoga, la observación de 200 test negativos de inmuno-deficiencia nos permitirá afirmar, por ejemplo, que la proporción de estudiantes seropositivos es *probablemente* menor del 0.5%; más precisamente, que la probabilidad de que tal proporción sea menor del 0.5% es 0.84. Finalmente, los resultados obtenidos en las pruebas de resistencia realizadas a las 5 piezas de un lote nos permitirán afirmar que la *probabilidad* de que una pieza del lote escogida al azar resista menos de 50 kg/cm² es de 0.018, por lo que podremos *predecir* que alrededor de un 1.8% de las piezas del lote resultarán inservibles.

El modelo probabilístico. Los *teoremas de representación* de la teoría de la probabilidad permiten *garantizar* que si los datos $D = \{x_1, \dots, x_n\}$ están efectivamente constituidos por un conjunto de observaciones homogéneas x_i , entonces *existe* un *modelo*, $p(x | \omega)$ que describe la *relación probabilística* entre los datos obtenidos y la naturaleza del proceso, descrita por un *parámetro o estado de la naturaleza* ω , (formalmente definido como límite de alguna función de las observaciones), de forma que D constituye una *muestra aleatoria* de tal modelo, y que *existe* además una distribución de probabilidad $p(\omega)$ sobre los posibles estados de la naturaleza que describe la información de que inicialmente se dispone sobre la naturaleza del proceso. El modelo probabilístico depende directamente de las condiciones experimentales. En problemas de investigación complejos, la determinación de un modelo probabilístico adecuado puede ser un problema muy difícil, que exija el trabajo en equipo de científicos que conozcan bien las características del problema y de matemáticos capaces de formalizarlas adecuadamente. Sin embargo, en problemas elementales, como la mayor parte de los que aparecen en los libros de texto, el modelo probabilístico puede ser directamente *deducido* del contexto experimental. Por ejemplo, en el caso de la tuberculina, las pruebas de laboratorio, realizadas tanto a personas sanas como a enfermos de tuberculosis, han permitido determinar que el test de la tuberculina resulta positivo en el 98% de los enfermos de tuberculosis, de forma que $p(+ | \text{Tuberculosis}) = 0.98$ (verdaderos positivos), y en el 5% de las personas sanas, de forma que $p(+ | \text{Sano}) = 0.05$ (falsos positivos). Por otra parte, en los casos de *medidas* repetidas de una misma magnitud, el modelo puede ser frecuentemente establecido a partir de la propia teoría de la probabilidad. Por ejemplo, si puede afirmarse que la variabilidad de las medidas realizadas es la consecuencia de muchas causas actuando independientemente unas de otras y sin que ninguna de ellas domine a las demás, entonces el *teorema central del límite* permite afirmar que el modelo probabilístico que las describe es *normal*; este sería el caso de las determinaciones de la constante gravitatoria o de las pruebas de resistencia mencionadas en los ejemplos anteriores, de forma que tales observaciones pueden ser analizadas como muestras aleatorias de poblaciones *normales*. En caso de los tests de inmunodeficiencia, la hipótesis de homogeneidad de la muestra directamente implica que se trata de observaciones dicotómicas (test positivo o negativo) con una probabilidad (constante, pero desconocida) de dar lugar a un resultado positivo, que es precisamente la proporción de estudiantes seropositivos. Es importante subrayar que la elección de un modelo probabilístico adecuado es una condición *indispensable* para la validez del análisis estadístico, cualquiera que sea el paradigma desde el que se analicen los datos. Sin embargo, el paradigma bayesiano presenta la enorme ventaja de disponer de los poderosos teoremas de representación para la difícil tarea de precisar tal modelo; por ejemplo, los resultados sobre homogeneidad parcial permiten construir y justificar los complejos *modelos jerárquicos* que recientemente han empezado a utilizarse, con notable éxito, en todo tipo de aplicaciones.

La distribución inicial. Los teoremas de representación *demuestran* la *necesidad lógica* de describir la información inicial de que se disponga sobre la naturaleza del proceso que

genera los datos, mediante una *distribución de probabilidad* $p(\omega)$ sobre los posibles estados de la naturaleza. Consecuentemente, si se dispone de información inicial *objetiva* es necesario *formalizarla* mediante probabilidades; en nuestro ejemplo de la tuberculina, se sabe que la proporción de enfermos de tuberculosis en la población estudiada es del 1%; consecuentemente, la distribución inicial de probabilidad en este caso es $\{p(\text{Tuberculosis}) = 0.01, p(\text{Sano}) = 0.99\}$. Frecuentemente, sin embargo, *no* se dispone de información inicial relevante, o la información inicial de que se dispone es de tipo *subjetivo*, y se desea limitar el estudio a las conclusiones *objetivas* que puedan ser deducidas basándose *exclusivamente* en el modelo probabilístico aceptado y en los datos experimentales D efectivamente obtenidos. En este caso, es necesario recurrir al *análisis de referencia*, que utiliza la *teoría de la información* para determinar la forma matemática de aquella distribución inicial $p(\omega)$ que describe una situación en la que *no* se dispone de información inicial sobre el verdadero valor de ω ; tal distribución recibe en nombre de *distribución inicial de referencia* y suele representarse mediante $\pi(\omega)$.

El proceso de inferencia. El *teorema de Bayes*, uno de los resultados básicos de la teoría de la probabilidad, permite *cuantificar* el proceso de inferencia, estableciendo la relación existente entre la distribución inicial $p(\omega)$, que describe la incertidumbre de la que se parte sobre el verdadero valor ω del estado de la naturaleza, y la *distribución final* $p(\omega | D)$ que describe la *incertidumbre residual* sobre el verdadero valor de ω , una vez observados y analizados los datos experimentales D . El teorema de Bayes tiene una expresión general muy sencilla,

$$p(\omega | D) \propto p(D | \omega) p(\omega),$$

donde \propto significa ‘proporcional a’, que especifica que la probabilidad final $p(\omega | D)$ asociada a cualquier estado de la naturaleza ω es *proporcional* a la probabilidad $p(D | \omega)$ de haber obtenido los datos D efectivamente observados si ese valor de ω hubiese sido el verdadero estado de la naturaleza, multiplicada por la probabilidad inicial $p(\omega)$ de tal valor. Los logaritmos de las probabilidades son medidas de información; consecuentemente, tomando logaritmos en la expresión anterior, el teorema de Bayes puede ser asimismo descrito afirmando que la *información final* de la que se dispone sobre el verdadero valor del estado de la naturaleza es la *suma* de la información de que inicialmente se parte, más la información que proporcionan los datos. El uso del teorema de Bayes es obviamente central en un sistema de inferencia basado en la teoría de la probabilidad; su uso continuado justifica el nombre de metodología *bayesiana* con el que generalmente se conoce este paradigma.

En la metodología bayesiana, la distribución final $p(\omega | D)$ de la magnitud de interés ω condicionada a los datos observados D es la solución al problema de inferencia planteado: resume *todas* las deducciones (necesariamente probabilísticas) que pueden hacerse sobre el valor de ω a la vista de los datos observados, el modelo probabilístico aceptado y la información inicial disponible. Si ω es una magnitud multidimensional, las conclusiones sobre cada una de sus componentes se obtienen calculando sus distribuciones *marginales*; por ejemplo, si $\omega = \{\mu, \sigma\}$, las conclusiones sobre μ que pueden deducirse de los datos D vendrán dadas por $p(\mu | D) = \int p(\omega | D) d\sigma$.

Cuando el número de estados posibles de la naturaleza, $\omega_1, \omega_2, \dots$, es finito (o infinito numerable), su distribución final, queda simplemente descrita por las probabilidades correspondientes, $\{p(\omega_1 | D), p(\omega_2 | D), \dots\}$, que obviamente deben sumar la unidad. En el ejemplo de la tuberculina, los análisis de laboratorio indican que $p(+ | \text{Tuberculosis}) = 0.98$ y que $p(+ | \text{Sano}) = 0.05$; además, la información inicial proporcionada por los estudios epidemiológicos indica que $p(\text{Tuberculosis}) = 0.01$; consecuentemente, por el teorema de Bayes,

$$\begin{aligned} p(\text{Tuberculosis} | +) &\propto p(+ | \text{Tuberculosis})p(\text{Tuberculosis}) \propto 0.98 \times 0.01 = 0.0098, \\ p(\text{Sano} | +) &\propto p(+ | \text{Sano})p(\text{Sano}) \propto 0.05 \times (1 - 0.01) = 0.0495; \end{aligned}$$

normalizando, la distribución final resulta ser

$$p(\text{Tuberculosis} | +) = \frac{0.0098}{0.0098 + 0.0495} = 0.165, \quad p(\text{Sano} | +) = \frac{0.0495}{0.0098 + 0.0495} = 0.835.$$

Observese que, contradiciendo la intuición de muchos, aunque el test de tuberculina tiene porcentajes de error relativamente pequeños (un 2% de falsos negativos y un 5% de falsos positivos), *únicamente* el 16.5% de las personas con test positivo de tuberculina estarán realmente enfermas de tuberculosis; naturalmente esto es una consecuencia de que la mayor parte de la población esta sana y, consecuentemente, la mayor parte de los positivos observados son falsos positivos; sin embargo, éste es un buen ejemplo de la utilidad de los métodos cuantitativos para controlar la intuición.

Cuando ω es una variable continua, su distribución final, $p(\omega | D)$ es descrita mediante una *densidad de probabilidad*, esto es, una función no negativa que encierra un área unidad, de forma que los estados de la naturaleza ω más verosímiles son los que corresponden a valores mayores de esta función, y que la probabilidad de que el verdadero estado de la naturaleza ω se sitúe entre dos valores cualesquiera a y b es $\int_a^b p(\omega | D) d\omega$, es decir el área encerrada por la función entre los valores a y b . Cuando los datos constituyen una muestra aleatoria de una población normal, y no se dispone de información inicial sobre sus parámetros (la media y la varianza de la población), la distribución final de referencia de la media poblacional (obtenida marginalizando el resultado de aplicar el teorema de Bayes), es una distribución de Student centrada en la media m de los datos, con parámetro de escala $s/\sqrt{n-1}$, donde s es la desviación típica de los datos, y con $n-1$ grados de libertad.

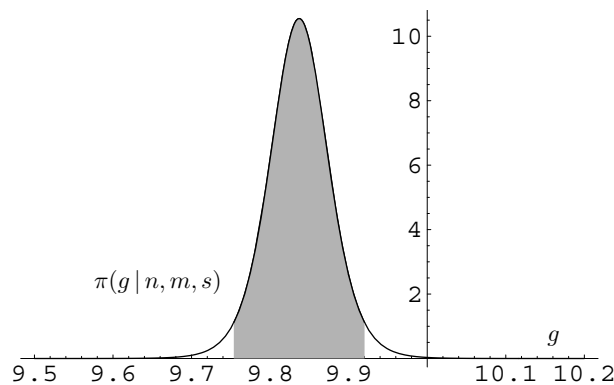


Figura 1. Conclusiones sobre el valor de la constante gravitatoria g basadas en un modelo normal y en los resultados experimentales.

En el ejemplo de las medidas de la constante gravitatoria g , la distribución *final* de referencia, correspondiente a una situación en la que no se supone información alguna sobre el verdadero valor de g ni sobre la precisión del método de medida, es la distribución de Student centrada en 9.837 m/seg^2 , parámetro de escala 0.037 , y 9 grados de libertad, cuya función de densidad de probabilidad está representada en la Figura 1. Como puede observarse en esa figura, las 10 medidas experimentales realizadas de la constante gravitatoria (relativamente imprecisas), permiten afirmar, prácticamente con certeza, que el verdadero valor de g se encuentra entre 9.7 y 10.0 (puesto que la densidad de probabilidad fuera de esa región es prácticamente cero). Además, la probabilidad de que, por ejemplo, el verdadero valor de g se sitúe entre 9.75 y 9.92 será el área limitada por la densidad final entre esos valores (la zona sombreada en la

figura), que resulta ser 0.95. Consecuentemente, la metodología bayesiana permite inferir de los datos experimentales que la probabilidad (grado de creencia) de que el verdadero valor de g esté dentro del intervalo $[9.75, 9.92]$ es 0.95. Esta afirmación puede compararse con el concepto convencional (frecuentista) de intervalo de confianza, que sólo permite afirmar que si se *repite el procedimiento* con muchas muestras, los intervalos correspondientes contendrían al verdadero valor de g en un 95% de los casos. Obviamente, sin embargo, esta última *no* es la cuestión relevante para el científico, que se pregunta por la probabilidad (grado de creencia) de que *en este caso concreto*, con los datos observados, el verdadero valor de g esté dentro del intervalo: ésta es la pregunta a la que contesta el paradigma bayesiano, afirmando que tal probabilidad es precisamente 0.95.

Cuando los datos están constituidos por un conjunto de observaciones dicotómicas independientes (se observa la existencia o la ausencia de una determinada característica en un conjunto de individuos elegidos al azar en una determinada población), y no se dispone de información inicial sobre la proporción θ de elementos en la población que poseen tal característica, la distribución final de referencia de θ , obtenida mediante el teorema de Bayes, es una distribución Beta con parámetros $r + \frac{1}{2}$ y $n - r + \frac{1}{2}$, donde r es en número de individuos que muestran la característica estudiada entre los n individuos observados.

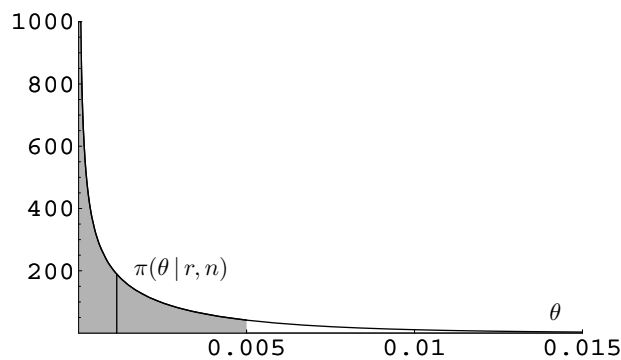


Figura 2. Conclusiones sobre la proporción θ de seropositivos entre los estudiantes basadas en el modelo binomial y en los resultados experimentales.

En el caso de la proporción θ , de estudiantes seropositivos, dada la información de que en una muestra aleatoria de $n = 200$ estudiantes no se ha observado ningún seropositivo ($r = 0$) la distribución *final* de referencia, correspondiente a una situación en la que no se supone información inicial alguna sobre el verdadero valor de θ , es la distribución Beta con parámetros $r + \frac{1}{2} = 0.5$ y $n - r + \frac{1}{2} = 200.5$, representada en la Figura 2. La densidad de probabilidad representada en esa figura contiene *toda* la información que los datos proporcionan sobre el valor de θ . Como puede observarse en la figura, esos 200 resultados experimentales permiten afirmar, prácticamente con certeza, que el verdadero valor de θ es menor de 0.015 (puesto que la densidad de probabilidad para valores de θ mayores es prácticamente cero). Además, la probabilidad de que el verdadero valor de θ sea, por ejemplo, menor que 0.005 resulta ser 0.843, que es el área de la parte sombreada, y la probabilidad de que el verdadero valor de θ sea menor que 0.0011 resulta ser 0.5 (y por lo tanto, la probabilidad de que sea *mayor* que 0.0011 es también 0.5). Si se requiere un *estimador* de θ , podríamos citar el valor 0.0011 (la mediana de la distribución final, representada por una línea en la figura) como una *estimación* de la probabilidad θ de que un estudiante cualquiera sea seropositivo, puesto que es igual de probable que θ sea mayor o menor que ese valor. El lector puede compararlo con la solución tradicional, el estimador

máximo-verosímil, que es la frecuencia relativa $\hat{\theta} = r/n$, y que resulta obviamente absurdo en este caso (cuando no se ha encontrado ningún seropositivo en la muestra el estimador máximo-verosímil es siempre igual a cero, cualquiera que sea el tamaño de la muestra observada). En cualquier caso, un estimador no es *nunca* una descripción suficiente de un resultado inferencial; una descripción *mínima* de las conclusiones que el paradigma bayesiano permite extraer de los datos, (descritas por la distribución final representada en la figura), sería decir que, a la vista de los resultados experimentales, la proporción de seropositivos es ciertamente menor del 1.5%, probablemente menor del 0.5%, y que es aproximadamente igual de probable que sea mayor que que sea menor del 0.1%.

La distribución predictiva. La distribución final $p(\omega | D)$ describe las conclusiones probabilísticas que pueden extraerse sobre el verdadero valor ω del estado de la naturaleza en base a los resultados experimentales D . Frecuentemente, sin embargo, el interés del investigador se centra en la incertidumbre asociada a futuras observaciones x del mismo proceso. Naturalmente, si ω fuese conocido, la incertidumbre sobre el valor de una observación futura estaría perfectamente descrita por la distribución $p(x | \omega)$ que especifica el modelo establecido; el verdadero estado de la naturaleza ω es desconocido pero, una vez observados los datos, tenemos sobre su valor la información proporcionada por $p(\omega | D)$. Consecuentemente, el *teorema de la probabilidad total* nos permite establecer que la información disponible sobre una observación futura puede ser expresada como la *media ponderada* correspondiente, esto es

$$p(x | D) = \int_{\Omega} p(x | \omega) p(\omega | D) d\omega.$$

Aplicando este resultado al análisis de los tests de inmunodeficiencia, la probabilidad de que un estudiante escogido al azar sea seropositivo (S) será $p(S | r, n) = \int p(S | \theta) p(\theta | r, n) d\theta$, lo que se reduce al valor medio de la distribución final de θ , que resulta ser $(r + \frac{1}{2}) / (n + 1) \approx 0.0025$. Consecuentemente, basándose únicamente en los 200 análisis negativos observados, la probabilidad (grado de creencia) de que un estudiante elegido al azar sea seropositivo es 0.0025.

Cuando los datos constituyen una muestra aleatoria de una población normal, y no se dispone de información inicial sobre sus parámetros, la distribución predictiva de referencia de una nueva observación, obtenida mediante el teorema de la probabilidad total, es una distribución de Student centrada en la media m de los datos, con parámetro de escala $s\sqrt{(n+1)/(n-1)}$, donde s es la desviación típica de los datos, y con $n - 1$ grados de libertad. En el caso de la resistencia de las piezas que constituyen un lote industrial, la distribución *predictiva* de referencia, correspondiente a una situación en la que *no* se supone información inicial alguna sobre valor medio de la resistencia de las piezas del lote ni sobre su variabilidad, es la distribución de Student centrada en la media, 60.98 kg/cm², de las 5 observaciones, con parámetro de escala 3.51, y con 4 grados de libertad, representada con trazo grueso en la Figura 3. Como puede observarse, las cinco observaciones realizadas permiten afirmar que la resistencia de cualquier pieza del lote estará seguramente entre 50 y 70 kg/cm²; específicamente, el área encerrada por la densidad predictiva entre esos límites es 0.951 y, consecuentemente, con la información disponible podemos esperar que el 95.1% de las piezas del lote se situarán en ese intervalo. Además, la probabilidad de que una observación futura sea menor de 50 kg/cm² es 0.018, que es el área de la parte sombreada; consecuentemente, puede esperarse que el 1.8% de las piezas del lote estarán por debajo de ese valor y que, por lo tanto, no cumplirán la especificación requerida.

Para predecir observaciones futuras los métodos estadísticos convencionales suelen recurrir a la distribución $p(x | \hat{\omega})$ que se obtiene substituyendo en el modelo probabilístico aceptado el

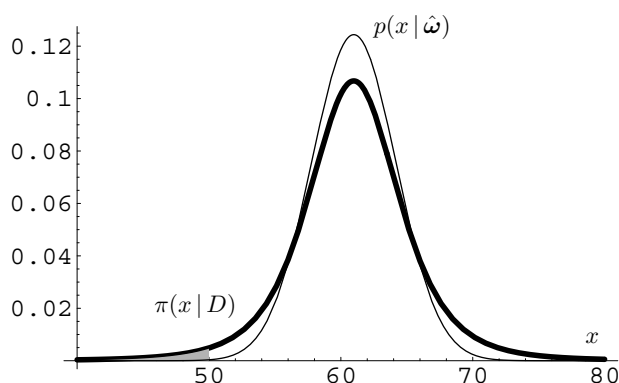


Figura 3. Conclusiones sobre la resistencia x de una nueva pieza basadas en el modelo normal y en los resultados experimentales.

verdadero valor ω del parámetro por un estimador suyo $\hat{\omega}$; esto significa trabajar como si $\hat{\omega}$ fuese el verdadero valor de ω , ignorando la incertidumbre existente sobre su verdadero valor. El resultado es una distribución de probabilidad *injustificadamente precisa* (salvo cuando las muestras son muy grandes, en cuyo caso el procedimiento convencional proporciona una *aproximación* razonable al resultado bayesiano exacto). En el ejemplo comentado se obtiene la densidad de probabilidad representada en la Figura 3 con trazo fino, según la cual la proporción estimada de piezas defectuosas sería del 0.03% (sesenta veces más pequeña de la estimación correcta, 1.8%). Como se trata de una afirmación probabilística sobre una magnitud *observable* (la presión soportada por nuevas piezas), es una afirmación *verificable*: el departamento de calidad de la empresa podría comprobar que, efectivamente, el porcentaje de piezas defectuosas en nuevas medidas se sitúa alrededor del 1.8%, no alrededor del 0.03% como los métodos convencionales sugerirían.

Algunas Referencias. En los últimos años han aparecido varios libros de texto sobre la metodología estadística bayesiana. En orden ascendente de nivel matemático, podemos recomendar: (i) D. A. Berry (1996), *Statistics, A Bayesian Perspective*, Duxbury Press; (ii) A. Gelman *et al.* (1995), *Bayesian Data Analysis*, Chapman & Hall; y (iii) J. M. Bernardo & A. F. M. Smith (1994), *Bayesian Theory*, Wiley. Los *Proceedings* del último congreso mundial de estadística bayesiana, celebrado en la Comunidad Valenciana hace unos meses, (Bernardo *et al.* eds., 1999, *Bayesian Statistics 6*, Oxford University Press), proporcionan una visión de conjunto sobre las últimas líneas de investigación y sobre los nuevos campos de aplicación del paradigma bayesiano.