TRES PARADIGMAS EN LA INFERENCIA ESTADÍSTICA

THREE PARADIGMS IN STATISTICAL INFERENCE

Ignacio **Méndez-Ramírez**[1†], Hortensia **Moreno-Macías**[2*], Chiharu **Murata**[3], Felipe de J. **Zaldívar-López**[4]

[1]Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, UNAM. Circuito escolar 3000, Ciudad Universitaria, 04510, Ciudad de México. [2]Universidad Autónoma Metropolitana, Unidad Iztapalapa. Av. San Rafael Atlixco 186, Leyes de Reforma primera sección, 09340, Ciudad de México. (hmm@xanum.uam.mx). [3]Instituto Nacional de Pediatría. Avenida de los Insurgentes Sur 3700, 04530, Ciudad de México. (chiharumurata@gmail.com). [4]Grupo de Atención Médica Integral. Pintapán sección 3 manzana 3 lote 3. Unidad Habitacional Cananea, Iztapalapa, 09969, Ciudad de México. (drzaldivar2@gmail.com).

INTRODUCCIÓN

En la naturaleza y en la sociedad, la construcción del conocimiento sobre la realidad se enfrenta a la variabilidad de los procesos estudiados. Una estrategia para tratar esa variabilidad es considerarla como un fenómeno aleatorio. En los fenómenos aleatorios no es posible hacer una predicción exacta o establecer leyes y teorías determinísticas acerca de algún evento. La estadística, para contender con la aleatoriedad se ha desarrollado como una disciplina que estudia la regularidad con la que ocurren los resultados de un fenómeno aleatorio al considerar grupos de elementos que comparten ciertas características, que delimitan a la población como concepto estadístico. El principio de regularidad estadística establece que las frecuencias relativas con las que aparecen los diferentes resultados, tenderán a cambiar muy poco conforme el número de observaciones aumenta, es decir esas frecuencias se estabilizan. Bajo el principio de regularidad estadística, los posibles resultados o variantes de un fenómeno aleatorio se pueden valorar a través de la frecuencia con que éstos ocurren. Los modelos matemáticos que representan las frecuencias estabilizadas de estos fenómenos se llaman distribuciones de probabilidad.

Las poblaciones se estudian mediante muestras de elementos que generan datos numéricos (Méndez-Ramírez *et al.*, 2013). La inferencia estadística es un proceso que trata de encontrar caracterizaciones de esa regularidad en las poblaciones a partir de los datos de las muestras. Existen tres paradigmas en la inferencia

*Autor responsable ❖ Author for correspondence.
Recibido: marzo, 2018. Aprobado: marzo, 2019.
Publicado como ENSAYO en Agrociencia 53: 1043-1069. 2019.

INTRODUCTION

In nature and in the society, the construction of knowledge on reality is fraught with variability of the processes studied. An strategy to deal with this variability is to consider it as a random phenomenon. In random phenomena, it is not possible to make an accurate prediction or to establish deterministic laws and theories about any event. Statistics to deal with randomness was developed as a discipline that studies the regularity with which the results of a random phenomenon occur within a group of elements that share certain characteristics delimiting the population as a statistical concept. The principle of statistical regularity establishes that the relative frequencies with which different results appear will tend to change very little as the number of observations increases, that is, those frequencies stabilize. Under this statistical regularity principle, the possible results or variants of a random phenomenon can be assessed using the frequency with which they occur. The mathematical models representing the stabilized frequencies of these phenomena are called probability distributions.

Populations are studied through samples of elements that generate numerical data (Méndez-Ramírez *et al.*, 2013). Statistical inference is a process that tries to find the characterization of the regularity in populations from samples data. There are three paradigms in statistical inference, but these are not clearly differentiated in the classroom neither in the common use of statistics. This paper aims to describe and compare the principal ideas of these three paradigms used in statistical inference, without delving into mathematical details. Firstly, the concepts of population and probability distributions are shown,

estadística, pero estos no están claramente diferenciados en la enseñanza y en el uso común de la estadística. En este escrito se pretende describir y comparar las ideas principales de los tres paradigmas usados en la inferencia estadística, sin pretender mostrar los detalles matemáticos. Inicialmente se presentan los conceptos de población y de distribución de probabilidad, y se ejemplifican algunos modelos comunes; después se describe cada paradigma y se comparan sus características.

Población y modelos de probabilidad

En la investigación y aun en la vida cotidiana se requiere contender con conjuntos de elementos ya sean personas, plantas, animales u objetos que constituyen las poblaciones de interés. Para estudiar casos prácticos, el conjunto de elementos que componen una población se acota, señalando las características que ese conjunto presenta. Por ejemplo, si lo que interesa es conocer la tasa de fecundidad en México durante el año 2020, la población a considerar está conformada por las mujeres que tienen entre 15 y 45 años y que viven en México durante ese año; si a cada mujer se le asigna el número de hijos que tiene (se mide) en ese periodo, se tendría un número para cada mujer de esa población, de tal manera que al promediar esos números se obtiene la llamada tasa de fecundidad. Evidentemente, al cambiar la población, por ejemplo, las mujeres con las mismas características, pero en diferente año de estudio, tendrán una tasa distinta.

Esas propiedades agregadas como promedios, proporciones, tasas, etc., caracterizan a la población, se llaman parámetros y se denotan comúnmente por letras del alfabeto griego, por ejemplo $\theta, \beta, \mu, \sigma, \lambda$.

En general, se establece que los elementos de la población son unidades u_i con ciertas características comunes denotadas por A, B, C, D. Así, con las ideas del ejemplo anterior podemos hablar de dos poblaciones: una conformada por mujeres con A) edad de 15 a 45, B) en México, C_1) en 2020; y otra por mujeres con A) edad de15 a 45, B) en México, C_2) en 2014. La medición del número de hijos en cada mujer es la variable de interés denotada comúnmente por X; y el agregado de esa variable de interés es la tasa de fecundidad θ_1 en la primera población y θ_2 en la segunda. Otro aspecto de interés es la determinación de la frecuencia de los diferentes valores de X; por

and some common models are exemplified; then, each of the paradigms are described and their characteristics are compared.

Population and probability models

In research and even in everyday life, it is required to deal with sets of elements, whether they are people, plants, animals or objects, that constitute the populations of interest. To study practical cases, the set of elements constituting a population is delimited, mapping out the characteristics that the set presents. For instance, if the interest is to know the fertility rate projection in Mexico during the year 2020, the population to consider will be compounded by women between 15 and 45 years old that will be dwelling in Mexico during that year. If each woman is assigned the number of children she has in that period, we will have a number for each woman of this population, and in such a way that by averaging these numbers, the so-called fertility rate projection will be obtained. However, if the population is changed, say, women with the same characteristics, but in different year of study; a different rate will be obtained.

These properties summed up as averages, proportions, rates, etc., that characterize the population are called "parameters", and are commonly denoted by letters of the Greek alphabet $\theta, \beta, \mu, \sigma, \lambda$, for example.

In general, it is established that the elements of the population are units u_i with certain common characteristics represented by A, B, C, and D. Thus, with the ideas of the previous example we can speak of two populations: one compounded by women with A) age 15 to 45, B) in Mexico, C_1) in 2020; and two by women with A) age 15 to 45, B) in Mexico, C_2) in 2014. The measurement of the number of children for each woman is the variable of interest, commonly denoted by X; and the aggregate of this variable of interest is the fertility rate θ_1 in the first population and θ_2 in the second. Another important aspect of interest is the determination of the frequency of the different X values. For instance, what is the frequency of women with zero children, or with 3 or 4 children, or with more than 4 children? To answer this, a model is used to represent the stabilized relative frequency of the occurrence of different values of X but that model and their

ejemplo, cual es la frecuencia de mujeres con cero hijos, o con 3 o 4 hijos, o con más de 4. Entonces, se usa un modelo para representar las frecuencias relativas estabilizadas de ocurrencia de los diversos valores de *X*, pero ese modelo y sus parámetros dependen de las características que definen a la población. Ese modelo es la llamada distribución de frecuencias o de probabilidades $f(x_i|\theta)$. En este caso los valores x_i cambian de un elemento a otro, pero el modelo pretende reflejar lo mejor posible las frecuencias con que ocurren los valores x_i en la población. Si esas frecuencias son estables se pueden considerar probabilidades de ocurrencia de los diversos valores de *X*. Estas probabilidades dependen de qué población se trata, qué elementos son y qué característica se les mide.

La elección de un modelo de distribución de probabilidad

La búsqueda del modelo que se ajusta mejor a la distribución poblacional de los datos, $f(x_i|\theta)$, se hace tomando en cuenta varios aspectos: la naturaleza del rango de valores de los datos, experiencias previas en poblaciones semejantes, consideraciones teóricas respecto a la constitución de la x_i, facilidad matemática y el grado de ajuste (bondad de ajuste) de la información de una muestra a un modelo en particular. Hay gráficas que permiten visualizar el grado de ajuste, como el histograma o las gráficas cuantil-cuantil en las que se observa el grado de concordancia de los cuantiles del modelo con los cuantiles empíricos en una muestra de elementos de la población. Hay pruebas estadísticas que permiten evaluar la bondad del ajuste, como la de Shapiro-Wilks, Kolmogorov-Smirnov, Anderson-Darling, etc. También hay estadísticas que permiten comparar el grado de ajuste: el criterio de información de Akaike (AIC) y el criterio bayesiano de información (BIC). Las distribuciones de probabilidad pueden ser discretas o continuas, según la naturaleza del rango de valores de los datos.

Distribuciones discretas

Algunos fenómenos tienen un número discreto de posibles resultados y es de interés contarlos. Por ejemplo, en una prueba clínica se cuenta el número de participantes que sanaron al recibir cierto tratamiento. También puede ser de interés contar el número de

parameters depend upon the characteristics that define the population. That model is the so-called frequency or probability distributions $f(x_i|\theta)$. In this case, the values of x_i change from one element to another, but the model tries to reflect, as much as possible, the frequencies with which the values of x_i occur in the population. If these frequencies are stable, probabilities of occurrence of the different *X* values can be considered. These probabilities depend on what population is being handled, on what are the existing elements, and on the characteristic being measured.

The choice of a probability distribution model

The search for a model that best fits the distribution of populational data, $f(x_i|\theta)$, is done taking into account several aspects such as: the nature of the data values range, previous experience in similar populations, theoretical considerations regarding the constitution of x_i, the mathematical facility and the degree of adjustment (goodness of fit) of the information from a sample to a particular model. There are graphs such as histogram or quantile-quantile that allows the visualization of adjustment degree, in which the extent of agreement of the quantiles of the model with the empirical quantiles in a sample of elements of the population can be observed. There are statistical tests employed to evaluate the goodness of adjustment, such as the test of Shapiro-Wilks, Kolmogorov-Smirnov, Anderson-Darling, etc. Besides, there are tests to compare the degree of adjustment: the statistics of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The probability distribution can be discrete or continuous depending on the nature of the range of data values.

Discrete distributions

Some phenomena may have a discrete number of possible outcomes and it is pertinent to count them up. For instance, in a clinical trial, the number of participants who were healed on receiving certain treatment should be counted. It may also be of interest to count the number of traffic accidents that occur in a fixed period of time. The binomial and Poisson distributions are two models that respectively represent such situations. It is worth mentioning that there are

accidentes de tránsito que ocurren en un periodo fijo de tiempo. Las distribuciones binomial y Poisson son dos modelos que representan respectivamente tales situaciones. Cabe mencionar que existen otros modelos para representar la regularidad de las variables discretas de acuerdo al contexto en el que ocurren.

Ensayos Bernoulli y distribuciones derivadas

Los ensayos de Bernoulli (1667-1748) son eventos o sucesos independientes unos de otros y con una probabilidad constante p de que ocurra cierta característica A que llamamos "éxito".

Distribución binomial

Si hay ensayos independientes, es de interés observar las frecuencias relativas con las que ocurren las muestras de n ensayos, con 0, 1, 2, 3,..., n éxitos. Así, si consideramos un grupo de cinco pacientes ($n=5$) que reciben tratamiento y la probabilidad de éxito (el paciente mejora) es θ; a esos pacientes los podemos considerar como cinco ensayos independientes de Bernoulli con probabilidad de éxito θ. *La regularidad con la que ocurre* cada posible número (n) de éxitos es conocida como distribución binomial con parámetros θ y n. Se puede demostrar que la probabilidad de que ocurran x éxitos, condicionada al valor de θ y de n es:

$$P(x\,|\,\theta)=C_x^n\theta^x(1-\theta)^{n-x} \text{ para } x=0, 1, 2, 3,\ldots, n \text{ y } 0\leq q\leq 1$$

El valor esperado y la varianza de esta distribución están dados por las siguientes expresiones:

$E(X)=n\theta$ y $Var(X)=n\theta(1-\theta)$ respectivamente.

En la Figura 1 se presentan tres distribuciones binomiales con diferentes parámetros (θ, n). Cabe notar que la forma es simétrica cuando $\theta=0.5$ como es el caso de la curva central, pero si $0\leq\theta<0.5$, la curva es sesgada a la derecha y sesgada a la izquierda cuando $0.5<\theta\leq 1$ como se observa en las otras curvas. Los valores esperados de las tres distribuciones son 3, 5 y 10.5, respectivamente (se observa un desplazamiento a la derecha) y las varianzas son 2.1, 2.5 y 3.15, en ese orden.

some models to represent the regularity of the discrete variables depending on the context in which they occur.

Bernoulli trials and derived distributions

Bernoulli trials (1667-1748) are events that are independent of each other with a constant probability p that certain characteristic A, called "success", occurs.

Binomial distribution

If independent trials are performed, it will be interesting to observe the relative frequencies with which the samples from n trials would occur with 0, 1, 2, 3,..., n successes. Thus, if we consider a group of five patients ($n=5$) that are receiving treatment and the probability of success (the patient improves) is θ; these patients can be considered as five independent Bernoulli trials with a probability of success θ. *The regularity with which* every possible number (n) of successes occur is known as binomial distribution with parameters θ and n. It can be shown that the probability of occurring successes, conditioned to the value of θ and of n is:

$$P(x\,|\,\theta)=C_x^n\theta^x(1-\theta)^{n-x} \text{ for } x=0, 1, 2, 3,\ldots, n \text{ and } 0\leq\theta\leq 1$$

The expected value and the variance of this distribution are given by the following expressions:

$E(X)=n\theta$ y $Var(X)=n\theta(1-\theta)$ respectively.

Figure 1 depicts three binomial distributions with different parameters (θ, n). It is worthwhile to note that the form is symmetric when $\theta=0.5$ as it the case of the central curve, but if $0\leq\theta<0.5$, the curve slants to the right and skewed to the left when $0.5<\theta\leq 1$ as it is observed the other curves. The expected values of the three distributions are 3, 5 and 10.5, respectively (a rightward shift is observed) and the variances are 2.1, 2.5 and 3.15 in this order.

Poisson distribution

Siméon Denis Poisson (1781-1840), a nineteenth-century French probabilist became interested in events

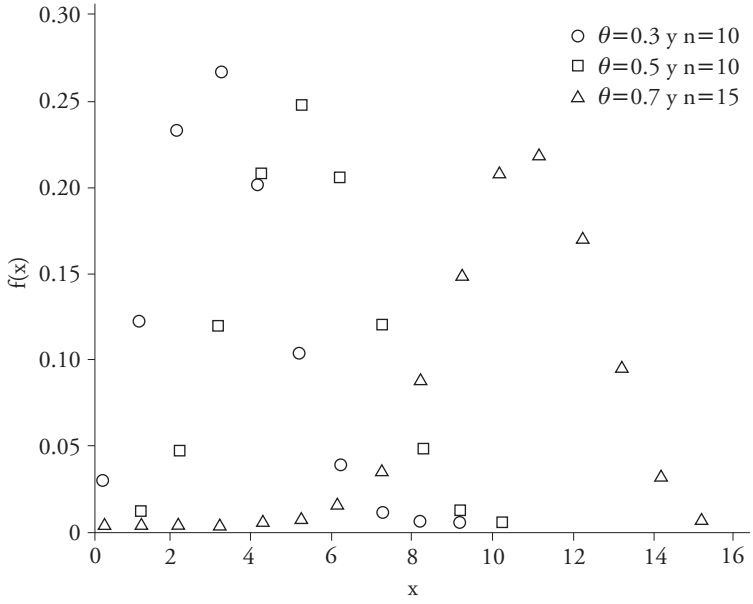

Figura 1. Distribuciones binomiales.
Figure 1. Binomial distributions.

Distribución Poisson

Siméon Denis Poisson (1781-1840), probabilista francés del siglo XIX, se interesó por los sucesos que se presentan de manera independiente y aleatoria en el transcurso del tiempo o el espacio. Suponiendo que el evento ocurre a una tasa constante, la función de probabilidad está dada por la expresión:

$$P(X=x\,|\,\lambda)=\frac{e^{-\lambda}\lambda^{x}}{x!} \text{ para } x=0,\ 1,\ 2,\ldots \text{ con } \lambda>0$$

donde: λ representa la tasa promedio de ocurrencia del evento por unidad de tiempo.

Un ejemplo famoso de la aplicación de la distribución Poisson es el hallazgo de L. Bortkiewics (1868-1931) quien estudió las frecuencias del número de muertes (X) por patada de caballo en el ejército Prusiano entre 1875 y 1894. Él observó que, pese a que las frecuencias variaban de un año a otro, y entre batallones, había un patrón de estabilización que se ajustaba a esta distribución.

La Figura 2 muestra tres distribuciones obtenidas con este modelo. Nótese que la forma de la distribución es asimétrica positiva pero conforme el valor de λ aumenta, la forma tiende a ser simétrica.

that occur independently and randomly in space or over time. Assuming that if an event occurs at a constant rate, the probability function is given by the expression:

$$P(X=x\,|\,\lambda)=\frac{e^{-\lambda}\lambda^{x}}{x!} \text{ for } x=0,\ 1,\ 2,\ldots \text{ with } \lambda>0$$

where, λ represents the average rate of occurrence of the event by unit of time.

A famous example of the application of Poisson distribution is the finding of L. Bortkiewicz (1868-1931). Bortkiewicz studied the frequencies of deaths (X) by horse kick in the Prussian army between 1875 and 1894. He observed that, despite the variations in the frequencies year to year and among battalions, there was a stabilization pattern that fitted this distribution.

Figure 2 shows three distributions obtained with this model. It should be noted that the distribution shape is asymmetrically positive, but as the value of λ increases, the shape tends to be symmetric.

In a binomial distribution, when n is large an θ is small, the binomial probability can be approximated through Poisson distribution. It can be shown that, in the limit, when $n\to\infty$ and $\theta\to 0$, the binomial

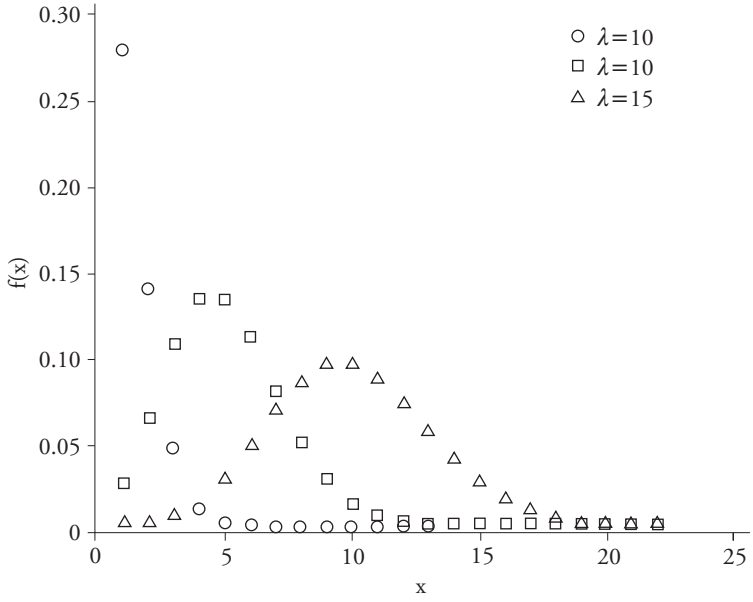

Figura 2. Distribuciones Poisson.
Figure 2. Poisson distributions.

Cuando en una distribución binomial n es grande y θ es pequeña, la probabilidad binomial puede aproximarse a través de la distribución Poisson. Se puede demostrar que, en el límite, cuando $n\to\infty$ y $\theta\to 0$, la distribución binomial converge a la Poisson con $\lambda = n\,\theta$ (siempre que λ tome un valor finito y sea constante). Esto es,

$$\lim_{\substack{n\to\infty \\ \theta\to 0}} P(x\,|\,\theta) = \lim_{\substack{n\to\infty \\ \theta\to 0}} C_x^n \theta^x (1-\theta)^{n-x}$$

$$= P(X = x\,|\,\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

donde: λ es el único parámetro de la distribución. También se puede demostrar que la varianza y el valor esperado de la distribución Poisson son ambos iguales a λ

Distribución binomial negativa

En el contexto de ensayos de Bernoulli se genera otra distribución, en la que, a diferencia de la distribución binomial, ahora la variable aleatoria cuenta el número de ensayos independientes n que se requieren para obtener exactamente x éxitos con una probabilidad de éxito por ensayo igual a θ. En esta distribución los parámetros son x y θ. El modelo es:

distribution converges to the Poisson with $\lambda = n\,\theta$ (provided that λ takes a finite value and is constant). That is:

$$\lim_{\substack{n\to\infty \\ \theta\to 0}} P(x\,|\,\theta) = \lim_{\substack{n\to\infty \\ \theta\to 0}} C_x^n \theta^x (1-\theta)^{n-x}$$

$$= P(X = x\,|\,\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

where, λ is the only parameter of the distribution. It can also be shown that the variance and the expected value of the Poisson distribution are both equal to λ.

Negative binomial distribution

In the context of Bernoulli trials, different distribution is generated, in which, unlike binomial distribution, the random variable considers the number of independent trials n required to get exactly x successes with a probability of success per trial being equal to θ. In this distribution, the parameters are x and θ. The model is:

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

$$n = 1, 2, 3, \ldots, x \text{ and } 0 \le \theta \le 1$$

$$P(n \mid x,\theta) = \binom{n-1}{x-1}\theta^{x}(1-\theta)^{n-x} \text{ para}$$

$$n=1, 2, 3,\ldots, x \text{ y } 0\leq\theta\leq 1$$

El valor esperado y la varianza están dadas por las siguientes expresiones:

$$E(X) = \frac{x(1-\theta)}{\theta}; \quad Var(X) = \frac{x(1-\theta)}{\theta^2}$$

En la Figura 3 se presentan dos distribuciones con diferentes parámetros

Distribuciones continuas

También hay fenómenos que tienen un número infinito de resultados posibles. Por ejemplo, aquellos que resultan de medir el tiempo de vida, o las características antropométricas (peso, estatura, circunferencia de cintura, etc.) de los miembros de una población. Dado que el número de posibles resultados es infinito, la probabilidad de un valor particular es cero, por lo que es necesario evaluar las probabilidades de intervalos de valores de la variable. Los modelos para estos casos se plantean en términos de funciones de densidad

The expected value and the variance are given by the following expressions:

$$E(X) = \frac{x(1-\theta)}{\theta}; \quad Var(X) = \frac{x(1-\theta)}{\theta^2}$$

In Figure 3, two distributions with different parameters are presented.

Continuous distributions

There are also phenomena that have an infinite number of possible results. For instance, those that result from measuring the time of life or anthropometric characteristics (weight, height, hip circumference, etc.) of the members of a population. Since the number of possible results is infinite, the probability of a particular value is zero; therefore, it is necessary to evaluate the probabilities of intervals of the values of the variables. The models for these types of cases are presented in terms of density function and the evaluation of probabilities corresponds to the area under the density curve in that interval. There are great varieties of models; perhaps, the most common examples are gamma, exponential and normal distributions.

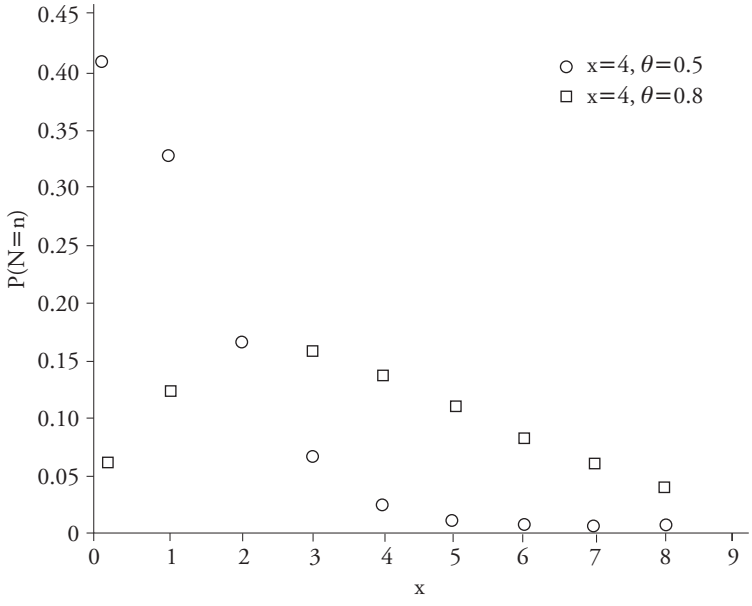

Figura 3. Distribución binomial negativa.
Figure 3. Negative binomial distribution

y la evaluación de las probabilidades corresponde al área bajo la curva de esa densidad en ese intervalo. Existe una gran diversidad de modelos, aunque quizá los ejemplos más comunes son las distribuciones gama, exponencial y normal.

Distribución gama

Aunque el número de años que vivirá una persona es impredecible, John Graunt (1620-1674) encontró frecuencias estables de longitud de vida en un conjunto de personas estudiado en Londres. Es obvio que la variable longitud de vida tiene valores de cero a un número grande y el modelo utilizado es:

$$P(a<X<b\,|\,k,\theta)=\int_a^b \frac{1}{\Gamma(k)\theta^k}x^{k-1}\exp\left(-\frac{x}{\theta}\right)dx$$

donde: $k<0$ y $\theta>0$ son los parámetros de forma y escala, respectivamente. La función está definida para $x>0$.

Este modelo tiene aplicación en diversas áreas y, en particular, se usa ampliamente para modelar tiempos de falla que ocurren de manera independiente a una tasa promedio. El valor esperado y la varianza son: $E(X)=k\theta$; $Var(X)=k\theta^2$. En la Figura 4 se ilustran dos funciones de densidad con diferentes parámetros: si $k=1$, se tiene una distribución exponencial; si $k>1$, la curva presenta un pico en $x=\theta(k-1)$.

Distribución exponencial

El tiempo que transcurre hasta que se presenta un evento particular cuya ocurrencia no depende del tiempo que ha transcurrido, se puede modelar usando una distribución exponencial. Por ejemplo, en las líneas de espera se mide el tiempo que transcurre entre un cliente y otro; o en el control de calidad de un producto se mide el tiempo en que se presenta una falla. La expresión matemática para esta distribución se deriva de la distribución gamma al hacer $k=1$.

$$P(a<X<b\,|\,\theta)=\int_a^b f(x)dx=\int_a^b \frac{1}{\theta}\exp\left(-\frac{x}{\theta}\right)dx$$

para $x>0$ y $\theta>0$

Gamma distribution

Although the number of years a person will live is unpredictable, John Graunt (1620-1674) in his study of human population in London found stable frequencies of the length of life in a set of people he studied in the city. Undoubtedly, the variable length of life have values from zero to a large number and the model used is:

$$P(a<X<b\,|\,k,\theta)=\int_a^b \frac{1}{\Gamma(k)\theta^k}x^{k-1}\exp\left(-\frac{x}{\theta}\right)dx$$

where, $k<0$ and $\theta>0$ are the parameters for shape and scale, respectively. The function is defined by $x>0$.

This model has application in various areas and, in particular, it is widely used to model failure times that occur independently at an average rate. The expected value and the variance are: $E(X)=k\theta$ and $Var(X)=k\theta^2$. In Figure 4, two density functions with different parameters are illustrated: if $k=1$, an exponential distribution is obtained; if $k>1$ the curve presents a peak in $x=\theta(k-1)$.

Exponential distribution

The time which elapsed before the presentation of a particular event whose occurrence does not depend on the time transpired can be modelled using exponential distribution. For example, in a queue, it is measured the time it takes between the turn of one client and the turn of another; or in quality control of a product, the time in which a failure presents is measured. The mathematical expression for this distribution is derived from gamma distribution by making $k=1$.

$$P(a<X<b\,|\,\theta)=\int_a^b f(x)dx=\int_a^b \frac{1}{\theta}\exp\left(-\frac{x}{\theta}\right)dx$$

for $x>0$ and $\theta>0$

As a consequence, $E(X)=\theta$ and $Var(X)=\theta^2$. Figure 5 is a graph of two exponential distributions with different parameters.

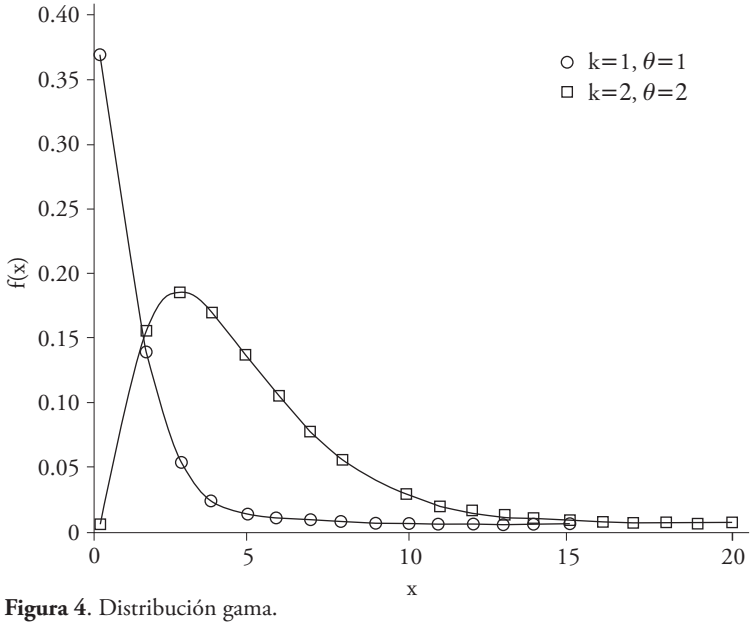

Figura 4. Distribución gama.
Figure 4. Gamma distribution

En consecuencia, $E(X)=\theta$ y $Var(X)=\theta^2$. La Figura 5 es una gráfica de dos distribuciones exponenciales con diferentes parámetros.

Distribución normal

Carl Friedrich Gauss (1777-1855) y Pierre Simon Laplace (1749-1827) postularon un modelo

Normal distribution

Carl Friedrich Gauss (1777-1855) and Pierre Simon Laplace (1749-1827) postulated a model called error distributions to represent the occurrence frequencies of the values of many measurements of a characteristic of the same object. The mean of those measurements was considered as the "true

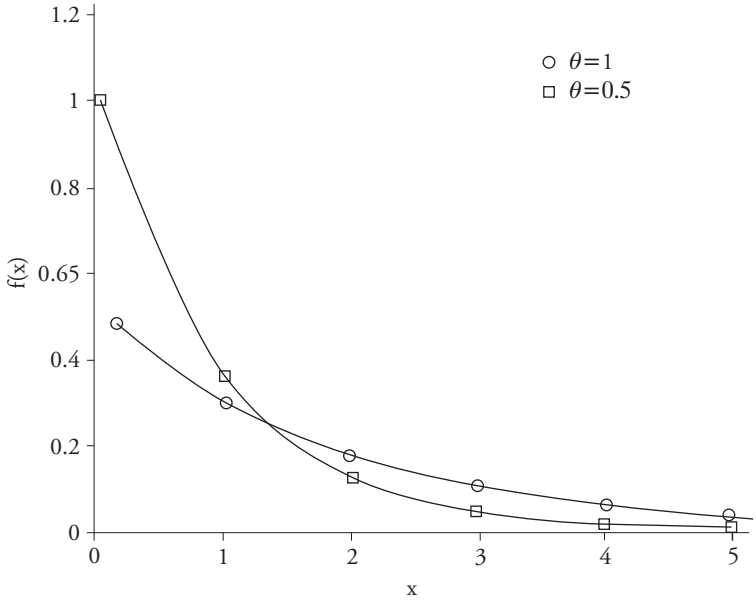

Figura 5. Distribuciones exponenciales.
Figure 5. Exponential distributions.

llamado distribución de errores para representar las frecuencias de ocurrencia de los valores de muchas mediciones de una característica de un mismo objeto. El promedio de esas mediciones se consideró el "valor verdadero" de esa característica. Lambert Adolphe Jacques Quetelet (1796-1874) aplicó ese modelo en una situación diferente; es decir, de medir un solo objeto muchas veces, el midió la misma característica una sola vez pero en muchos sujetos. Él observó que esos valores varían alrededor de un valor llamado media, que para el caso de mediciones antropométricas lo llamó el "hombre medio", y que sus frecuencias se distribuyen con el mismo modelo de la distribución de errores que denominó "distribución normal".

El modelo expresa la probabilidad de que una variable aleatoria X tome valores en un intervalo (a, b) y depende de dos parámetros que son la media μ y la varianza σ^2. Los valores posibles para x y μ van de $-\infty$ a $+\infty$; los de σ van de 0 a ∞.

La expresión matemática del modelo normal con media μ y varianza σ^2 es la siguiente:

$$P\left(a<X<b\,|\,\mu,\sigma^2\right)=\int_a^b f(x)\,dx = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}dx$$

y se denota como $X\sim N(\mu, \sigma^2)$.

Hay un caso muy particular cuando $\mu=0$ y $\sigma^2=1$ denotada como $Z\sim N(0,1)$. Esta es la normal estándar que sirve como referencia para el cálculo de probabilidades en las distribuciones normales con cualquier par de parámetros después de aplicar la transformación (llamada estandarización): $z=\frac{x-\mu}{\sigma}$. La representación gráfica de la normal estándar está en la Figura 6. Una característica particular de esta distribución es que $P(-1<Z<1)\approx 0.68$; $P(-1.96<Z<1.96)\approx 0.95$; $P(-2.58<Z<2.58)\approx 0.99$.

Varias distribuciones normales se presentan en la Figura 7. Nótese que conforme aumenta la media, las curvas se desplazan a la derecha y en la medida que la varianza aumenta, la curva es más amplia (los datos están más dispersos).

value" of that characteristic. Lambert Adolphe Jacques Quetelet (1796-1874) applied that model in a different situation; that is, instead of measuring a characteristic many times in a single object, he measured the same characteristic only once in many subjects. He observed that these values vary around a value called average or mean, which in the case of anthropometric measurements he called it the "average man", and that their frequencies are distributed with the same model of error distribution that he called "normal distribution".

The model expresses the probability that a random variable X takes values in an interval (a, b) and depends on two parameters – the mean m and the variance σ^2. The possible values for x and m range from $-\infty$ to $+\infty$; and those of σ range from 0 to ∞.

The mathematical expression for normal model with an average m and variance σ^2 is as follows:

$$P\left(a<X<b\,|\,\mu,\sigma^2\right)=\int_a^b f(x)\,dx = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}dx$$

this is denoted as $X\sim N(\mu, \sigma^2)$.

There is a very particular case when $\mu=0$ and $\sigma^2=1$ denoted as $Z\sim N(0,1)$. This is the normal standard that serves as a reference for calculating probabilities in normal distributions with any pair of parameters after applying the transformation (called standardization): $z=\frac{x-\mu}{\sigma}$. Figure 6 shows the graphical representation of normal standard. A particular characteristic of this distribution is that $P(-1<Z<1)\approx 0.68$; $P(-1.96<Z<1.96)\approx 0.95$; $P(-2.58<Z<2.58)\approx 0.99$.

Several normal distributions are presented in Figure 7. Note that as the mean increases, the curves shift to the right and as the variance increases, the curves become wider (*i.e.* the data are more dispersed).

As mentioned, wherever a study of the society and nature is focused, there will be randomness, variability and unpredictability. However, due to the constancy in the frequencies of occurrence of the possible results, regularity is also found.

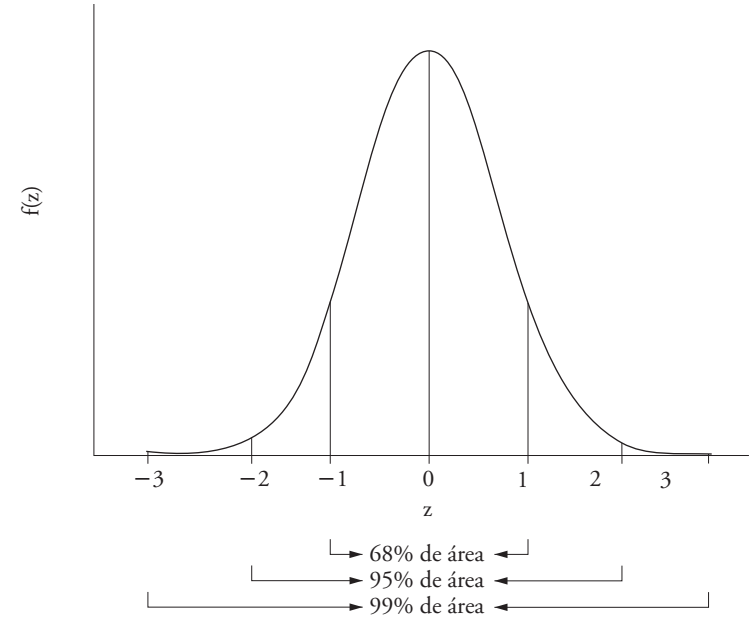

Figura 6. Distribución normal.
Figure 6. Normal distribution.

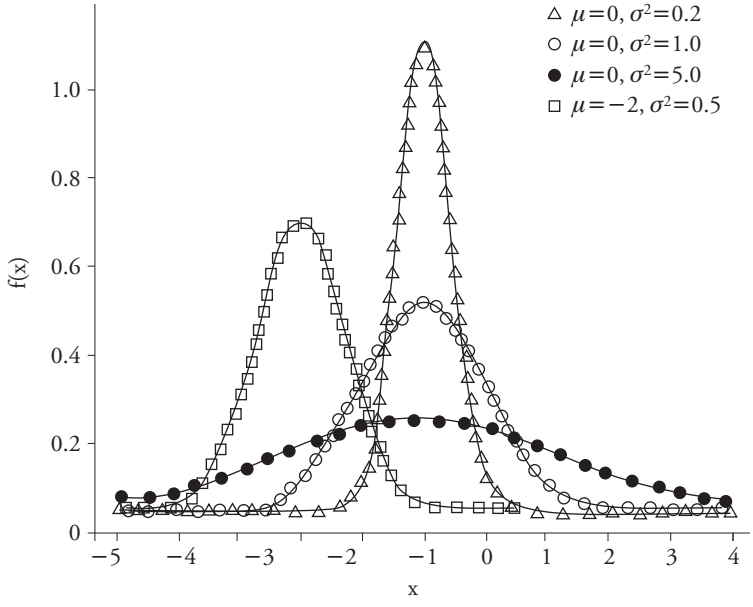

Figura 7. Distribuciones normales.
Figure 7. Normal distributions.

Como se mencionó, donde quiera que se enfoque un estudio de la sociedad y la naturaleza, se encontrará aleatoriedad, variabilidad e impredictibilidad. Sin embargo, debido a la constancia en las frecuencias de ocurrencia de los posibles resultados, también se encuentra regularidad.

Applied statistics characterizes and compares populations. If you use in this task models such as those already exposed (among others), it is called parametric statistics because it involves the value of the parameters. The probability distribution models are assumed to be known by their forms; however, they

La estadística aplicada caracteriza y compara poblaciones. Si en esta tarea utiliza modelos como los ya expuestos (entre otros), se le llama estadística paramétrica porque involucra el valor de los parámetros. Los modelos de distribución de probabilidades se suponen conocidos en su forma, pero dependen de parámetros para una especificación completa. La inferencia estadística paramétrica es una herramienta que permite valorar características de la población a partir de una muestra. Al considerar un modelo, los parámetros se estiman y las hipótesis relativas a ellos se contrastan.

La estimación inicia con un modelo y se exploran valores plausibles del parámetro, por lo cual es de tipo exploratorio. En la prueba de hipótesis se plantea, como pregunta de investigación, si se considera que los datos disponibles son compatibles o no con una cierta hipótesis sobre los valores del parámetro, por lo que se le puede considerar de tipo confirmatorio.

Paradigmas básicos de la inferencia estadística

De acuerdo con Chatterjee (2003) y Thompson (2007) hay tres paradigmas en la inferencia estadística: 1) comportamental; 2) en la instancia; y 3) Bayesiano. En los tres se parte de bases conceptuales diferentes e inconmensurables en el sentido de los paradigmas de Khun (1996). Por lo tanto, aunque el desarrollo matemático puede conducir a expresiones iguales o al menos muy parecidas, las interpretaciones son muy distintas. No obstante, con mucha frecuencia en la práctica de la estadística las diferencias son poco claras y motivan discusiones irreconciliables (Salsburg, 2001). En esta sección se pretende, de manera muy breve, caracterizar dichos paradigmas y enfatizar sus diferencias en concepción e interpretación.

En la mayoría de las situaciones, y para los tres paradigmas, se supone una función $f(x|\theta)$ que modela la aleatoriedad de los resultados. El o los valores de la variable aleatoria están representados por x (un número o un conjunto de números) y θ es el o los valores de los parámetros del modelo (un número o un conjunto de números). El modelo $f(x|\theta)$ se conoce como función de probabilidad en el caso discreto, y como función de densidad en el caso continuo.

depend on parameters for a complete specification. Parametric statistical inference is a tool that allows the assessment of the population characteristics from a sample. When considering a model, the parameters are estimated and the hypothesis related to them are contrasted.

Estimation begins with a model and then the plausible parametric values are explored; for this, it is an exploratory type. In the test of hypothesis, it is raised, as a research question, if the available data are considered compatible or no with a certain hypothesis on the parametric values; thus, it can be considered as confirmatory type.

Basic paradigms of statistical inference

According to Chatterjee (2003) and Thompson (2007), there are three paradigms in statistical inference: 1) behavioral, 2) in the instance and 3) Bayesian. The three paradigms are based on different and immeasurable conceptual bases of Khun paradigms (1996). Therefore, even when mathematical development can lead to equal or at least very similar expressions, the interpretations are very different. Nevertheless, very often in statistical practice, the differences are unclear and provoke irreconcilable discussions (Salsburg, 2001). In this section, it is intended, in a nutshell, to characterize these paradigms and highlight their differences in conception and interpretation.

In most situations and for the three paradigms, a function $f(x|\theta)$ that models the randomness of the results are assumed. The random variable value or values are represented by x (a number or a set of numbers) and θ is the value or values of the model parameters (a number or a set of numbers). The model $f(x|\theta)$ is known as probability function in the discrete case and as density function in the continuous case.

Compartment paradigm (Neyman, Pearson and Wald)

From research process perspective, this paradigm is identified before conducting the study. This means that there is an intention to establish a plan of analysis with it, but at the moment, the empirical or observed data are not yet generated.

Paradigma comportamental (Neyman, Pearson y Wald)

Desde el punto de vista del proceso de investigación, este paradigma se ubica antes de realizar el estudio. A través de él se pretende establecer un plan de análisis, pero aún no se tienen datos empíricos (observados).

Este paradigma se caracteriza por ser del tipo de toma de decisiones, en el que se considera a la estadística como un juego contra la naturaleza. Para cada decisión del estadístico y situación de la naturaleza hay una pérdida. Así se buscan estrategias o reglas de decisión que minimicen el promedio de las pérdidas. Todas están basadas en la conceptualización de tomar muchas muestras y en ellas el cálculo de una estadística, es decir, de una función de los datos que puede ser un estimador o una estadística de prueba, por ejemplo, la media, la proporción y la varianza como estimadores; el valor de t o de F como estadísticas de prueba. Éstas también se conocen como procedimientos "frecuentistas" porque se basan en la frecuencia de ocurrencia de valores de la estadística, al considerar que se toman, en teoría, todas las muestras posibles. Así se generan las llamadas "distribuciones derivadas del muestreo".

Estrategia de intervalo

La estrategia de estimación bajo este paradigma trata de construir intervalos de valores de los parámetros a partir de los estimadores puntuales. Se buscan intervalos que tengan la menor dimensión pero la mayor probabilidad de cubrir el "verdadero" valor del parámetro. Bajo este enfoque se supone que el parámetro es desconocido pero tiene un valor fijo.

Para cada muestra se obtiene un estimador puntual y un intervalo, de tal manera que si este proceso se repitiera un número grande de veces, se producirían intervalos que, digamos, en el 95% de los casos contienen al parámetro. Esto es, se tendría un intervalo al 95% de confianza. Por ejemplo, considere una población de hombres de 70 años, de los cuales un 35% (valor de θ) sobrevive hasta los 80 años. Al estimar este parámetro se construye un intervalo a partir de cada muestra. Tres de los posibles intervalos obtenidos pueden ser: (30-40), (34-44) o (36-46). En los dos primeros se cubre el valor del parámetro pero en el tercero no. Con frecuencia se interpretan erróneamente los intervalos de confianza cuando se señala que un intervalo en

This paradigm is decision-making type in which statistics is considered as a game against nature. For every decision of the statistician and natural situation, there is a loss. Thus, what is needed are strategies or decision rules that would minimize the average loss. The basis for all of them is the concept of taking many samples and on them the calculation of a statistic, that is, finding a function of the data that can be an estimator or a test statistic, for example, the mean, the proportion and the variance; and the value of t and F as a statistic test. These strategies are also known as "frequentist" procedures because they are based on the frequencies of occurrence of the statistical values on considering that all the possible samples are, in theory, included. In this way, the so-called "distributions derived sampling" are generated.

Interval strategy

The estimation strategy under this paradigm tries to construct intervals of parameter values from the point estimators. Intervals that have the smallest dimension but the largest probability to cover the "true" value of the parameter are looked for. Under this approach, it is assumed that the parameter is unknown but has a fixed value.

A point estimate and an interval are obtained for each sample, in such a way that if this process is repeated a large number of times, intervals that contain the parameter, say, in 95% of the cases would be obtained. That is to say, a 95% confidence interval would be obtained. For instance, consider a population of 70-year-old men, of which 30% (value of θ) survives up to 80 years. When estimating this parameter, an interval is constructed from each sample. Three of the possible intervals that can be obtained may be: (30-40), (34-44) or (36-46). In the first two, the value of the parameter is covered, but not in the third. Most often, the confidence intervals are erroneously interpreted when a particular interval is shown to cover or uncover the parameter, without minding that the process is what generates the intervals. The process is that with 95% of probability of producing an interval that covers the parameter value. In addition, it is pertinent to clarify that the researcher has no need to carry out the repetitions of the samples because it is performed based on the theoretical distribution of the estimator.

particular cubre o no cubre el parámetro, ignorando el proceso que genera los intervalos. El proceso es el que tiene el 95% de probabilidad de producir un intervalo que cubra el valor del parámetro. Además, es importante aclarar que el investigador no necesita realizar las repeticiones de las muestras porque se efectúa con base en la distribución teórica del estimador.

Considere un sujeto para el cual específicamente pasan los 10 años, ese sujeto en particular no muere o bien muere antes de los 80. Y si muere, "se muere todito", y no el 35% de él. Así, un "intervalo de confianza al 95% para un parámetro", contendrá ese parámetro en el 95% de los intervalos que se construyesen con las muestras posibles. Sin embargo, un intervalo obtenido para una muestra en particular, incluye o no incluye el parámetro. La aseveración es que proviene de un proceso en el que el 95% de ellos contiene al parámetro. Es como un reportero que dice "según fuentes generalmente bien informadas, ocurrió tal evento".

Estrategia de pruebas de hipótesis

En el proceso de "pruebas de hipótesis", se plantean dos hipótesis sobre el estado de naturaleza: nula (H_0) y la alternativa (H_a) que especifican cierto(s) modelo(s) de probabilidad considerados hipótesis nula (H_0), y otro(s) modelo(s) como hipótesis alternativa (H_a). Al inicio hay preferencia por la hipótesis nula; se diseñan reglas de decisión que según sea la muestra se rechaza la hipótesis nula o no se rechaza. Dos tipos de errores se pueden cometer: el error tipo 1 que consiste en rechazar la hipótesis nula cuando "la naturaleza" es esa hipótesis; el error tipo 2 si no se rechaza la hipótesis nula cuando la naturaleza presenta la alternativa. Se buscan procedimientos o reglas de decisión que fijan la probabilidad del error tipo 1 en un valor bajo, usualmente 0.05, y buscar que la probabilidad de cometer el error tipo 2 sea lo más pequeña posible mientras más alejado de la hipótesis nula esté el verdadero modelo (naturaleza) de generación de las variables aleatorias resumida en la estadística de prueba. Estas dos probabilidades de error se definen al considerar muchas posibles muestras aleatorias, que originan muchos valores de la estadística de prueba. Una regla o función de decisión se diseña y para esto, primero con los datos de cada muestra posible se obtiene el valor de una estadística de prueba; luego, el rango de valores de la estadística de prueba (llamado

Consider a subject for which specifically the 10 years passed and this particular subject does not die or dies before the age of 80. And if he dies, his death will be "total", and not just 35%. Thus, a "95% confidence interval for a parameter" will contain that parameter in 95% of the intervals that are constructed with the possible samples. However, an interval constructed for a particular sample include or does not include the parameter. The assertion is that it comes from a process in which 95% of them contain the parameter. This can be likened to a reporter who says "according to generally well-informed sources, such event occurred".

Hypothesis testing strategy

In "hypothesis testing" process, two hypotheses - the null (H_0) and the alternative (H_a) that specify certain probability model(s) - about the state of nature are proposed. Initially, there is a preference for the null hypothesis; decision rules are designed and depending on the sample the null hypothesis is rejected or not rejected. Two types of errors can be committed: the type 1 error consists in rejecting the null hypothesis when "the nature" supports the hypothesis; the type 2 error is not rejecting the null hypothesis when the nature presents the alternative. It is necessary to seek for procedures or decision rules that fix type 1 error probability at a low value, usually at 0.05, and try to reduce the probability of committing type 2 error as small as possible, while farther the null hypothesis is from the true model (nature) for generating the random variables summarized in the test statistics. These two error probabilities are defined by considering many possible random samples that give raise to many values of the test statistic. A rule or decision function is designed and for this, the first thing is to start with the data of each possible sample to obtain the value of a test statistic; then, the range of values of the test statistic, called rejection zone, that has a given probability (*vgr.* 0.05) to occur if H_0 is true. It is sought that this range of values of the test statistic has the minimum probability of occurring if the H_0 is not true and the truth is an alternative hypothesis H_a; this is the type 2 error. That is to say, the test is powerful; power defined as the probability of rejecting H_0 when it is false. This can be likened to a physician who diagnoses patients, and on seeing many patients, he obtains success in a high percentage of them (say 95%). This strategy has a wide field of application

zona de rechazo) que tiene una probabilidad dada (*vgr*. 0.05), de ocurrir si la H_0 es cierta. Se busca que ese rango de valores de la estadística de prueba tenga la mínima probabilidad de ocurrir si la H_0 no es cierta y lo cierto es una hipótesis alternativa H_a, éste es el error tipo 2. Es decir, que tenga alta potencia definida como la probabilidad de rechazar H_0 cuando ésta es falsa. Es como un médico que diagnóstica a los pacientes y al considerar muchos pacientes, tiene éxito en un porcentaje elevado de ellos (digamos 95%). Esta estrategia tiene un campo de aplicación muy amplio en diversas áreas tales como la epidemiología, la de seguros y la industrial entre otras; dado que en esos casos las posibles muestras se generan al repetir el proceso muchas veces. Su uso en investigaciones en las que la muestra se da por obtenida es cuestionable; como se mencionará más adelante.

Paradigma en la instancia (Fisher)

Desde el punto de vista del proceso de investigación, este paradigma se ubica después de que se tiene sólo una muestra (ya se tienen los datos); no interesan los valores de los datos en otras muestras.

La muestra se toma por una única vez, es decir, los datos están dados, así ocurrieron, de modo que se pregunta "qué modelo de probabilidad es el que hace que esos datos sean los más probables de ocurrir". Las ideas principales de este enfoque se ejemplificarán en el contexto de n ensayos de Bernoulli. Si el modelo que produce los datos es conocido (una distribución binomial), pero no se conoce su parámetro θ; entonces la probabilidad de que ocurran x éxitos en n ensayos está dada por la función de probabilidad:

$$f(x|\theta) = P(x \text{ éxitos en } n \text{ ensayos dado } \theta) = C_x^n \theta^x (1-\theta)^{n-x}$$

En esta función, la condición se puede plantear al revés, esto es, como la función de θ condicional a la muestra x y así se obtiene otra función que se llama verosimilitud. La función de verosimilitud se denota usualmente como $\mathcal{L}(\theta|x)$.

$$\mathcal{L}(\theta|x) = p_\theta(x) = P_\theta(X=x)$$

Cabe aclarar que ésta no es función de probabilidad para θ; es la probabilidad con la cual cada valor de θ produce la muestra x.

in various areas such as epidemiology, insurances and in industries among others, since in these cases, the possible samples are generated on repeating the process many times. Its use in investigation in which the samples are considered obtained is questionable, as will be shown later.

Paradigm in the instance (Fisher)

From the point of view of the research process, this paradigm is identified after obtaining only a sample whose data are already available, without minding the values of the data in other samples.

The sample is taken only once, that is, the data are given, and they had happened. So, the question is "what probability model is what makes those data to be the most likely to occur?" We will exemplify the principal ideas of this approach in the context of n Bernoulli trials. If the model that produces the data is known (a binomial distribution), but its parameter θ is unknown; then the probability that x successes occur in n trials is given by the probability function:

$$f(x|\theta) = P(x \text{ successes in } n \text{ trials given } \theta) = C_x^n \theta^x (1-\theta)^{n-x}$$

In this function, the condition can be raised in the reverse way, that is, as the function of θ conditional to the sample x and in this way another function called "likelihood" is obtained. The likelihood function is usually denoted as $\mathcal{L}(\theta|x)$.

$$\mathcal{L}(\theta|x) = p_\theta(x) = P_\theta(X=x)$$

It is clear that this is not probability function for θ. It is the probability with which each value of θ produces the sample x.

The likelihood is the function of the parameters conditional to the given data. This expression gives the probability of the only sample x that occured (in the instance) according to the values of θ. The value of θ that maximizes the likelihood function is known as "maximum likelihood estimator" and is usually denoted with $\hat{\theta}$.

One way to represent the likelihood of various values of θ given that an x result occurred is to evaluate the relative likelihood or the quotient of each value of the likelihood divided by the evaluated likelihood in

La verosimilitud es función de los parámetros, condicional a los datos dados. Esta expresión da la probabilidad de la única muestra x que ocurrió (en la instancia) según los valores de θ. El valor de θ que maximiza la función de verosimilitud se conoce como "estimador de máxima verosimilitud" y usualmente se denota con $\hat{\theta}$.

Una forma de representar la verosimilitud de los diversos valores de θ dado que ocurrió un resultado x, es evaluar la verosimilitud relativa o el cociente de cada valor de la verosimilitud dividido entre la verosimilitud evaluada en su máximo valor. Esto se conoce como razón de verosimilitudes (RV). La expresión es:

$$RV=\mathcal{L}(\theta|x)/\mathcal{L}(\hat{\theta}|x)$$

Esta expresión se considera como función de θ cuando el valor de x es el que ocurrió en la instancia estudiada. Valores de θ cercanos a $\hat{\theta}$ también tienen una probabilidad "grande" de haber producido la muestra x. Cabe la pregunta ¿Qué es "grande"? La idea es que no disminuya mucho el valor de la RV. Así se dice que aquellos valores de θ que no disminuyen la verosimilitud "por mucho" son un intervalo de verosimilitud. Los valores de θ son los que hacen verosímil el valor de x que se tiene. Así, no se sabe cuál es el valor del parámetro θ, y entonces, se estima éste de manera que la verosimilitud sea la máxima (igualando la primer derivada de la verosimilitud a cero y despejando a θ como una función de x).

Por ejemplo, si se considera que una disminución de la verosimilitud desde 1 hasta 0.1 determina valores de θ que aún son factibles de haber producido la muestra x. Este conjunto de valores se representa con $\{\theta:(\theta|x)/(\hat{\theta}|x)\geq 0.1\}$. Es decir, el conjunto de valores de θ con razones de verosimilitud mayores o iguales a 0.1. En general, se puede construir un intervalo de verosimilitud, es decir, el conjunto de valores de θ que tienen una razón de verosimilitud mayor que un porcentaje dado son los que disminuyen la verosimilitud en ese porcentaje o menos.

$$\{\theta:\mathcal{L}(\theta|x)/\mathcal{L}(\hat{\theta}|x)\geq p/100\}$$

Nótese que este valor $p/100$ no es una probabilidad, sólo es un límite o cota para el apoyo relativo de los valores de θ comparados con el valor $\hat{\theta}$. A partir de una analogía de los intervalos de confianza al 95

its maximum value. This is known as likelihood ratio (LR). The expression is:

$$LR=\mathcal{L}(\theta|x)/\mathcal{L}(\hat{\theta}|x)$$

This expression is considered as the function of θ when the value of x is what occured in the instance studied. Values of θ close to $\hat{\theta}$ also have a "big" probability of having produced the sample x. Now, the question is "What is "big"?" The idea is do not have a large decrease in the RV value. Thus, it is assumed that those values of θ that do not diminish the LR "by far" are a range of likelihood. The values of θ are those that make credible the value of x. Thus, it is not known what is the value of the parameter θ, and so, this is estimated in such a way that the likelihood is the maximum (equalizing the first likelihood derivative to zero and clearing θ as a function of x).

For instance, let us assume that a decrease in the likelihood from 1 until 0.1 would determine the values of θ that are still possible to produce the sample x. This set of values is represented with $\{\theta:(\theta|x)/(\hat{\theta}|x)\geq 0.1\}$. That is, the set of values of θ with likelihood ratios greater or equal to 0.1. In general, it is possible to construct a range of likelihood, that is, a set of values of θ that have likelihood ratio greater than a given percentage, are the ones that decrease the likelihood in this given percent or less.

$$\{\theta:\mathcal{L}(\theta|x)/\mathcal{L}(\hat{\theta}|x)\geq p/100\}$$

Note that the value $p/100$ is not a probability but just a limit or level for the relative support for the values of θ compared with the value of $\hat{\theta}$. Based on the analogy of the 95 and 99% confidence intervals, Pawitan (2001) proposed two possible limits for the value $p/100$: which are 0.15 and 0.04.

It is common that instead of LR as a quotient of degrees of support, $-\log(\text{LR})$ is used as the measurement of the likelihood of each value of θ compared with that of $\hat{\theta}$. If the value of θ (which does not detract in its probability of producing x in relation to $\hat{\theta}$) is considered to be the one where LR is greater than 0.1, this is equivalent to the value of θ in which $-\log(\text{LR})$ is less than 2.3 (Figure 8).

y 99%, Pawitan (2001) propuso dos posibles límites para el valor de $p/100$: 0.15 y 0.04.

Es frecuente que en lugar de la RV como cociente de grados de apoyo se use $-\log(\text{RV})$ como medida de la verosimilitud de cada valor de θ comparada con la de $\hat{\theta}$. Si se considera que un valor de θ (que no desmerece en su probabilidad de producir x con relación a $\hat{\theta}$) es aquel donde la RV sea mayor que 0.1, esto es equivalente a el valor de θ en que $-\log(\text{RV})$ sea menor que 2.3 (Figura 8).

En las pruebas de significancia se supone un valor o valores del parámetro $\theta=\theta_0$, llamada la hipótesis de nulidad dada x. Entonces, el supuesto de que H_0 es cierta sólo se rechaza si el valor de máxima verosimilitud representa un grado de contradicción "grande" entre θ_0 y $\hat{\theta}$. El máximo de la función se obtiene bajo el supuesto de que la hipótesis nula es cierta $P(x|\theta_0)$. Este grado de contradicción se determina de nuevo con una razón de verosimilitud, que ahora es:

$$RV = \frac{P\left(x \mid \hat{\theta}_0\right)}{P\left(x \mid \hat{\theta}\right)}$$

Si esta RV es pequeña, digamos inferior a 0.1, se rechaza H_0, y decimos que los valores de θ en H_0 entran en contradicción, o sea, son mucho menos verosímiles que el estimador de máxima verosimilitud. Si esto no sucede, no se rechaza H_0. El límite para el rechazo

In significance tests, a value or values of the parameter $\theta=\theta_0$ called null hypothesis when x given is assumed. Therefore, the supposition that H_0 is true is only rejected if the maximum likelihood value represents a "big" degree of contradiction between θ_0 and $\hat{\theta}$. The maximum of the function is obtained under the supposition that the null hypothesis $P(x|\theta_0)$ is true. This degree of contradiction is newly determined with a likelihood ratio which is now:

$$LR = \frac{P\left(x \mid \hat{\theta}_0\right)}{P\left(x \mid \hat{\theta}\right)}$$

If LR is small, say less than 0.1, H_0 is rejected, and we say the values of θ in H_0 are in contradiction, that is, they are likely much less credible than the maximum likelihood estimator. However, if this does not happen, H_0 is not rejected. The limit for the rejection is determined when the value of θ that satisfices the hypothesis distracts or produces a much lower probability. Then, the only x observed without restriction is the value of $\hat{\theta}$, and for this reason, the hypothesis is rejected.

The probability P is determined based on the fact that if H_0 is true, x data are produced with the degree of contradiction observed with the hypothesis or even greater. The value of P is the degree of contradiction between the null hypothesis and the specific value of

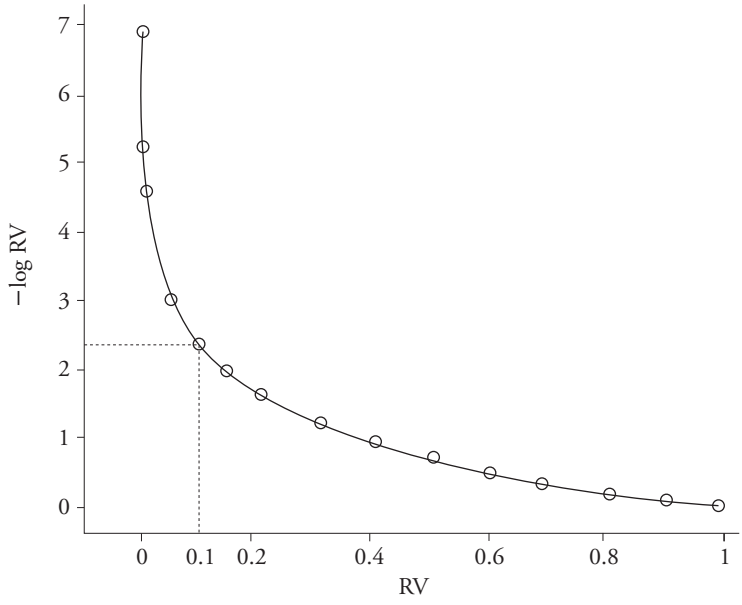

Figura 8. Razón de verosimilitud.
Figure 8. Likelihood ratio.

se determina cuando el valor de θ que cumple con la hipótesis desmerece o produce una mucho menor probabilidad, la única *x* observada es el valor de $\hat{\theta}$ sin restricciones, por lo que se rechaza la hipótesis.

Se determina la probabilidad *P* de que si H_0 es cierta, se produzcan datos *x* con el grado de contradicción observado con la hipótesis, o aún mayor. El valor de *P* es el grado de contradicción entre la hipótesis de nulidad y el valor específico de los datos en la muestra. Esta es una decisión semejante a lo expresado por Popper (1959), se rechaza una hipótesis H_0 si *P* es "pequeña"; pero si *P* no es "pequeña", no se rechaza la hipótesis H_0 aunque no se demuestra que sea verdadera, sólo queda como parte de las hipótesis plausibles.

Lo que hizo Fisher fue una extensión del silogismo *modus tollen* de la lógica. El silogismo expresa: si *H* es cierta, entonces debe ocurrir cierto resultado *q*, pero si en la práctica no ocurre *q*, se rechaza *H*. Veamos un ejemplo sencillo. Suponga que la hipótesis *H* es que llovió en la noche, con el supuesto adicional que se tiene un lugar descubierto. Si llueve, se moja el suelo en ese lugar. Aquí *q* es que el suelo está mojado. Si al observar ese sitio al día siguiente está seco, se rechaza la *H*, es decir, se establece que no llovió. A esto también le han llamado la *negación del consecuente*. En este ejemplo se ve que el rechazo ocurre con certeza; en cambio, la aceptación de *H* no es válida. En el ejemplo anterior, si el sitio observado está mojado, no es seguro que llovió, pudo regarse, o se rompió una tubería, o hubo una inundación, como hipótesis alternativas. En el postulado de Fisher se trata de lo mismo, pero ahora hay aleatoriedad. Entonces, el consecuente de *H* es algo que ocurre con probabilidad alta. Si esto no ocurre se tiene una disyuntiva; *H* no es cierta o bien, *H* es cierta pero ocurrió un evento improbable. Para valorar qué tan improbable es ese evento, se obtiene el valor de P. Este valor es la probabilidad de que ocurra un alejamiento o contradicción de los datos con la *H* como el que se presentó o aún mayor. El investigador usa este valor como una indicación del grado de contradicción de los datos con la hipótesis. No se requiere tomar la decisión de rechazar *H,* mucho menos aceptarla. Este enfoque está más ligado a la investigación científica que a un proceso productivo, natural o social que se repite. En este último caso el paradigma comportamental es el adecuado.

Es importante observar que para la evaluación del valor de *P* se requiere conocer la distribución de la estadística de prueba, es decir, la distribución derivada

the data in the sample. This is a decision similar to that expressed by Popper (1959): a H_0 hypothesis is rejected if *P* is "small", but if *P* is not "small", the H_0 hypothesis is not rejected although it is not proven to be true; it remains only as part of the plausible hypothesis.

What Fisher did was an extension of the *modus tollen* syllogism of the logic. The syllogism expresses: if *H* is true, then a certain result *q* must occur, but if in practice, *q* did not occur, *H* is rejected. Let us look at a simple example. Assume that the hypothesis *H* is that it rained at night, with an additional assumption that there was uncovered place. If it rained, the soil in that place would get wet. Here, *q* is that the ground is wet. If this place was observed the next day and the ground was dry, *H* is rejected, that is, it is established that it did not rain. This has also been called "*negation of the consequent"*. In this example, it is seen that the rejection occurs with certainty; however, the acceptance of *H* is not valid. In the previous example, if the site observed is wet, it is not all sure that it rained as the site could be watered or a pipe might be broken or that there was a flooding. All these can be alternative hypotheses. Fisher's postulation is the same; however in this time, there is randomness. Therefore, the consequent of *H* is something that happens with a high probability. If this does not happen, there is a dilemma. That is *H* is not true or well, *H* is true but an unlikely event occurred. To evaluate how unlikely is that event, the value of *P* is obtained. This value is the probability that a shift or contradiction of the data occur with *H* as the one presented or even greater. The researcher uses this value as an indication of the degree of contradiction of the data with the hypothesis. It is not necessary to make the decision to reject *H*, much less to accept it. This approach is more linked to scientific research than to a productive, natural or social process that is repeated. In this last case, the behavioral paradigm is appropriate.

It is important to observe that for the evaluation of the *P* value, knowledge of test statistic distribution is required, that is, the sample derived distribution. Nevertheless, a rejection zone is not established in the values of this statistic *a priori*. It is used only to obtain the value of *P*.

The model can represent the expected situation, the negation of a causality hypothesis. The model is the hypothesis to test which is usually called H_0. The

del muestreo. Sin embargo, no se establece una zona de rechazo en los valores de esta estadística *a priori*; se usa solo para obtener el valor de *P*.

El modelo puede representar la situación esperada, la negación de una hipótesis de causalidad. El modelo es la hipótesis a probar usualmente llamado H_0. La concordancia entre los datos observados (*O*) y los esperados (*E*) se evalúa bajo el supuesto que el modelo es cierto. Usualmente esta discordancia se valora de acuerdo con la probabilidad (valor de *P*) de una discordancia como la obtenida o aún mayor, suponiendo cierto el modelo dado por H_0. Por ejemplo, si el modelo postula un valor de cero para la diferencia entre dos promedios poblacionales $\mu_1-\mu_2=0$, se espera que el valor de *t* de Student sea 0 (el valor esperado). Una vez que se hace la investigación y se tiene un valor especifico de la estadística *t* de Student (t_0), se valora el grado de alejamiento de 0 que tiene ese valor t_0 o un alejamiento aun mayor $P(t>t_0)$ ese es el "valor de *P*". La interpretación más común, dentro de este paradigma es:

$P<0.01$ — Evidencia fuerte en contra de H_0
$0.01<P<0.05$ — Evidencia moderada en contra de H_0
$0.05<P<0.10$ — Evidencia sugestiva en contra de H_0
$P>0.10$ — Poca o nula evidencia en contra de H_0

Estas evaluaciones se usan rutinariamente en los artículos de investigación. Es importante notar que en el razonamiento anterior no se contempla el error tipo 2 ni una hipótesis alternativa particular, tampoco hay una región de rechazo. Sólo se pretende saber el grado de apoyo que la única muestra que se tiene da a la H_0.

Considere una prueba de diagnóstico que puede ser positiva (+) o negativa (−) y que esto depende de si el sujeto al que se le aplica la prueba está enfermo de un padecimiento específico (E) o no lo está (no E). Consideramos la tabla de probabilidades del resultado positivo (+) o negativo (−), según si es E o no E, como se muestra en el Cuadro 1.

concordance among the observed (*O*) data and the expected (*E*) data is evaluated under the assumption that the model is true. Usually, this discordance is evaluated based on the probability (value of *P*) of a discordance like the one obtained or even greater, with the assumption that the model given by H_0 is true. For instance, if the model postulates zero value for the difference between two population averages $\mu_1-\mu_2=0$, it is expected that the Student's *t*-value be 0 (the expected value). Once the research is done and there is specific value of the Student's *t*-statistic (t_0), the degree of shift from 0 that has the value t_0 or even greater shift $P(t>t_0)$ is evaluated and that is the "*P value*". The most common interpretation within this paradigm is:

$P<0.01$ — Strong evidence against H_0
$0.01<P<0.05$ — Moderate evidence against H_0
$0.05<P<0.10$ — Suggestive evidence against H_0
$P>0.10$ — Little or no evidence against H_0

These evaluations are routinely used in research articles. It is worthwhile to note that in the previous reasoning, type 2 error is not contemplated, nor a particular alternative hypothesis, nor is there a rejection region. It is only intended to know the degree of support that the only available sample gives.

Consider a diagnostic test that can be positive (+) or negative (−) and that this depends on whether the subject to whom the test is applied is sick of a specific disease (S) or is not (NS). Let's consider the probabilities table of positive result (+) or negative (−) depending on whether it is S or no S as shown in Table 1.

According to the paradigm in the instance, if a particular patient has a positive test, the likelihood ratio is calculated with the assumption that he is sick until proved otherwise. For some authors, this evidence is sufficiently small if the following quotient is less than 0.10 (Chatterjee, 2003).

Cuadro 1. Situación real contra los resultados de la prueba diagnóstica.
Table 1. Real situation against the results of the diagnostic test.

Resultado de la prueba	Enfermo (E)	No enfermo (NE)
Prueba positiva (+)	a	b
Prueba negativa (−)	c	d

De acuerdo con el paradigma en la instancia, si un paciente en particular tiene una prueba positiva, se calcula el cociente de verosimilitudes con el supuesto de que está enfermo hasta que se encuentre evidencia de que no lo está. Para algunos autores esta evidencia es suficientemente pequeña, si el siguiente cociente es menor de 0.10 (Chatterjee, 2003).

$$\frac{P(+|\,no\ E)}{P(+|\,E)} < 0.1$$

En muchos problemas estándar, pero no en todos, la inferencia en la instancia coincide numéricamente con el comportamental. Sin embargo, la interpretación y los supuestos que la sustentan son diferentes.

Paradigma bayesiano

A diferencia de los enfoques anteriores en los que los parámetros son constantes desconocidas, en el paradigma bayesiano los parámetros se consideran variables aleatorias en dos sentidos. Por un lado, a los parámetros fijos se les asocia una distribución de probabilidades que refleja el grado de incertidumbre que los investigadores tienen sobre sus valores. Pero existe la posibilidad de que el parámetro sea aleatorio bajo un contexto determinado. Los modelos de efectos aleatorios (intercepto y pendientes) son un ejemplo de esta situación. En ambos casos, la distribución es llamada "*a priori*" o "inicial".

Este paradigma opera una vez que se tienen los datos (x), al igual que en el paradigma en la instancia. También utiliza la función de verosimilitud como punto de partida. El Teorema de Bayes se usa para que con la distribución *a priori* y la verosimilitud se obtenga la probabilidad *a posteriori* o final, que es la distribución del parámetro condicionada al valor de x, es decir, dados los datos. Con la distribución final se obtienen conclusiones sobre el fenómeno estudiado. En este paradigma la verosimilitud se interpreta como una probabilidad condicionada de dos variables aleatorias x y θ:

$$P(x,\theta) = P(\theta)P(x|\theta) = P(x) = P(\theta|x) \Rightarrow\ P(\theta,x) = P(\theta)\ P(x|\theta)/P(x)$$

Este teorema se puede ver como una ponderación de la verosimilitud, donde el ponderador es:

$$\frac{P(+|\,no\ E)}{P(+|\,E)} < 0.1$$

In many standard problems, but not all, the inference in the instance coincides numerically with the compartmental one. Nevertheless, the interpretation and the assumptions that support it are different.

Bayesian paradigm

Unlike the previous approaches in which the parameters are unknown constants, in the Bayesian paradigm, the parameters are considered random variables in two directions. On one hand, the fixed parameters are associated with a distribution of probabilities that reflects the degree of uncertainty which the researchers have about their values. But there is a possibility that the parameter is random under a given context. The models of random effects (intercept and gradient) are an example of this situation. In both cases, the distribution is called "*a priori*" or "*initial*".

This paradigm operates once you have the data (x) as in the instance paradigm. It also uses the likelihood function as a starting point. The Bayes theorem is used so that with the *a priori* distribution and the likelihood, the *posteriori* or final probability is obtained, which is the distribution of the parameter conditioned to the value of x, which is the given data. With the final distribution, conclusions on the studied phenomenon are obtained. In this paradigm, the likelihood is interpreted as a conditioned probability of the two variable randoms x and θ:

$$P(x,\theta) = P(\theta)P(x|\theta) = P(x) = P(\theta|x) \Rightarrow\ P(\theta,x) = P(\theta)\ P(x|\theta)/P(x)$$

This theorem can be seen as a likelihood weighing where the weighting is:

$$P(\theta)/P(x)$$

This $P(\theta)$ represents what is known before the study on θ considering it as random.

The inference about the value of the parameter is done by obtaining the mean or mode of the final distribution. The range of probability *a posteriori* can

$$P(\theta)/P(x)$$

Esta $P(\theta)$ representa lo que se sabe antes del estudio sobre θ considerándola como aleatoria.

La inferencia sobre el valor del parámetro se hace obteniendo la media o moda de la distribución final. También se puede obtener un intervalo de probabilidad *a posteriori* al seleccionar un intervalo de valores de θ con una probabilidad alta, digamos del 95%, de que q esté en ese intervalo llamado intervalo de probabilidad *a posteriori*.

En la prueba de hipótesis bayesiana se valora la probabilidad de que el parámetro esté en el espacio de la hipótesis nula o en el espacio de la hipótesis alternativa. En el caso en que la hipótesis de nulidad especifique un solo valor de θ (hipótesis simple: H_0: $\theta=\theta_0$), se usa una distribución con cierta probabilidad para ese valor y el resto de la probabilidad se asigna a todos los demás valores de θ. En ambos casos se evalúa el cambio de la probabilidad inicial a la final. Si el cambio es "grande", se rechaza la hipótesis nula.

El enfoque bayesiano se puede ver como un proceso continuo de aprendizaje al usar una distribución final como la inicial en el siguiente paso. Con este proceso se mejora la distribución *a priori* de tal manera que se refleja cada vez más cercanamente a la realidad. Es como un médico que va modificando su distribución *a priori* al obtener más datos sobre un paciente y puede ofrecer un diagnóstico más certero en la medida que conoce más a ese paciente.

En las pruebas de tamizaje se puede considerar que cada valor de θ corresponde a cada individuo, la distribución inicial modela la prevalencia de los valores de θ. Cada individuo genera los datos de manera aleatoria.

El enfoque bayesiano puede producir resultados que numéricamente coinciden con los otros dos (comportamental y en la instancia); sin embargo, la interpretación es completamente diferente entre los tres. En la Figura 9 se ilustran las diferencias básicas en los supuestos de los tres paradigmas; los tres utilizan como punto de partida la función de verosimilitud.

Comparación de los tres paradigmas en el contexto de pruebas diagnósticas

Enfoque de Neyman-Pearson. En el Cuadro 1, el estatus de enfermedad es fijo pero desconocido y el resultado de la prueba es aleatorio. Equivale

also be obtained by selecting a range of θ values with a high probability, say 95%, of which θ is within this interval called *a posteriori* probability interval.

In Bayesian hypothesis test the probability that the parameter is within the space of null hypothesis or in the space of the alternative hypothesis is evaluated. In the case where the null hypothesis specifies only single value of θ (simple hypothesis: H_0: $\theta=\theta_0$), a distribution with certain probability for that value is used and the rest of the probability is assigned to all other values of θ. In both cases, the change from the initial probability to the final is evaluated. If the change is "big", the null hypothesis is rejected.

The Bayesian approach can be seen as a continuous learning process by using a final distribution as the initial one in the next step. With this process, the distribution *a priori* is improved in such a way that the reality is more and more closely reflected. It is like a doctor who is modifying his distribution *a priori* by getting more data on a patient and can offer a more accurate diagnosis as he knows the patient more.

In screening test, it can be considered that each value of θ corresponds to each individual and that the initial distribution models the prevalence of θ values. Each individual generates the data randomly.

The Bayesian approach can produce results that numerically coincide with behavioral and in the instance approaches; nevertheless, the interpretation is completely different among the three. In Figure 9, the basic differences in the assumptions of the three paradigms are illustrated. The three use likelihood function as a starting point.

Comparison of the three paradigms in the context of diagnostic test

Neyman-Pearson approach. In Table 1, the state of illness is fix but unknown, and the result of the test is random. It is equivalent to considering two samples of $a+c$ size for patients with that illness and the other of $b+d$ for the individuals without the illness. In both, the result of the test is classified. The research question can be made as follows: among the different tests, which is the best? To answer this, the following quotients should be calculated: $\text{sensibilidad} = P(+ \mid E) = \dfrac{a}{a+c}$

$$\text{especificidad} = P(- \mid NE) = \frac{d}{b+d}.$$

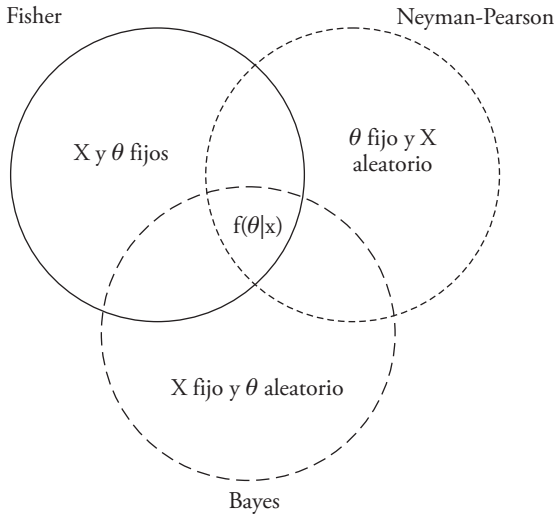

Figura 9. Diferencias básicas en los supuestos de los tres paradigmas.
Figure 9. Basic differences in the assumption of the three paradigms.

a considerar dos muestras de tamaño $a+c$ para pacientes con la enfermedad y la otra de tamaño $b+d$ para los individuos no enfermos; en ambas se clasifica el resultado de la prueba. Se puede plantear la pregunta de investigación: entre varias pruebas ¿cuál es la mejor? Para esto se calculan los siguientes cocientes: $\text{sensibilidad} = P(+|E) = \frac{a}{a+c}$;

$\text{especificidad} = P(-|NE) = \frac{d}{b+d}$.

En general se buscan pruebas más sensibles que específicas por considerar que es más importante identificar de manera correcta a los enfermos. Si se muestrea de manera global $a+b+c+d$ individuos, y éstos se clasifican de acuerdo con el Cuadro 1, entonces es válido obtener:

Valor Predictivo Positivo

$$(\text{VPP}) = P(E|+) = \frac{a}{a+b};$$

Valor Predictivo Negativo

$$(\text{VPN}) = P(NE|-) = \frac{d}{c+d}$$

como la probabilidad de estar (no estar) enfermo dado que la prueba sale positiva (negativa).

In general, more sensitive than specific tests are sought for because it is more important to identify correctly the patients. If $a+b+c+d$ individuals are sampled globally, and these are classified according to Table 1, then it is valid to obtain:

Positive Predictive Value (PPV)

$$(\text{PPV}) = P(E|+) = \frac{a}{a+b};.$$

Negative Predictive Value (NPV)

$$(\text{NPV}) = P(NE|-) = \frac{d}{c+d}$$

as the probability of being (not being) sick since the test is positive (negative).

If the result of the test is continuous, for instance, the creatinine or hemoglobin, it can also be asked: What is the optimum cutoff limit of each test? To answer this question, a graph of *sensitivity* against 1-*specificity* at different cutoff values is constructed. This graph is known as *Receiver Operating Characteristic* (ROC) curve, which allows to determine that cutoff point with maximum sensitivity and minimum 1-specificity.

In this paradigm, all the possible test results and all disease status are considered.

Si el resultado de la prueba es continua, por ejemplo, la creatinina o hemoglobina, también se puede preguntar: ¿Cuál es el límite de corte óptimo de cada prueba? Para contestarla se construye una gráfica de la *sensibilidad* contra 1-*especificidad* a diferentes valores de corte. Esta gráfica se conoce como "curva ROC (sigla en inglés de *Receiver Operating Characteristic*)" la cual permite determinar aquel punto de corte con máxima sensibilidad y mínima 1-especificidad.

En este paradigma se consideran todos los posibles resultados de la prueba y todos los estatus de la enfermedad.

Enfoque de Fisher. Se parte de una pregunta como la siguiente: si se usó una prueba específica y resultó positiva para un sujeto, ¿qué evidencia se tiene para sustentar un diagnóstico en ese paciente? La posibilidad de un resultado negativo no es de interés, porque no ocurrió. En el Cuadro 1, el estatus de enfermedad no es una variable aleatoria, sólo lo es el resultado de la prueba. Se valora $\frac{P(+|NE)}{P(+|E)}$ como la verosimilitud de enfermedad dado que la prueba salió positiva. Es decir, se busca la evidencia de que esté enfermo hasta que se demuestre lo contrario. Si el cociente anterior es "pequeño" (digamos, menor a 0.1), se favorece el diagnóstico de enfermedad. De manera semejante se puede tratar la situación cuando la prueba resultó negativa, y se valora entonces el cociente $\frac{P(-|NE)}{P(-|E)}$.

La pregunta es ¿cómo cambia la información *a priori* después de hacer la prueba?

Enfoque Bayesiano. En el Cuadro 1, el estatus de la enfermedad y el resultado de la prueba son aleatorios. Es decir, se muestrean $n=a+b+c+d$ individuos y con ellos se construye el cuadro.

La distribución *a priori* de la enfermedad se estima con los totales marginales del cuadro.

$$P(E)=\frac{a+c}{a+b+c+d}$$

Esta es la prevalencia de la enfermedad estimada con la muestra global ($a+b+c+d$). De manera semejante, entonces, la 1-*prevalencia de enfermedad* se estima con:

$$P(NE)=\frac{b+d}{a+b+c+d}$$

Fisher's approach. This starts with a question as follows: If a specific test was used and it resulted positive for a subject, "what evidence is available to support the diagnosis in that patient?" The possibility of a negative result is not of interest, because this did not occur. In Table 1, the disease status is not a random variable; only the result of the test. $\frac{P(+|NE)}{P(+|E)}$ is evaluated as the likelihood of illness given that the test is positive. That is, the evidence that the subject is ill is sought until proven otherwise. If the last quotient is "small" (say less than 0.1), diagnosis of the disease is favored. Similarly, the situation can be handled in the same way if the result is negative. Therefore, the quotient $\frac{P(-|NE)}{P(-|E)}$ is evaluated.

The question is "how does the information change *a priori* after doing the test?"

Bayesian approach. In Table 1, the disease status and the result of the test are randoms. That is, $n=a+b+c+d$ individuals are sampled and the table is constructed with them.

The *a priori* distribution of the disease is estimated with the marginal totals of the table.

$$P(E)=\frac{a+c}{a+b+c+d}$$

This is the prevalence of the estimated disease with the global sample ($a+b+c+d$). Similarly, the *disease* 1-*prevalence* is estimated with:

$$P(NE)=\frac{b+d}{a+b+c+d}$$

Assuming the test was positive for this patient, then the probability of the sickness, given that the test was positive, can be calculated with Bayes formula:

$$P(E|+)=\frac{P(E\cap+)P(E)}{P(+)}$$

$$=\frac{\left(\frac{a}{a+b+c+d}\right)\left(\frac{a+c}{a+b+c+d}\right)}{\frac{a+b}{a+b+c+d}}$$

$$=\frac{a(a+c)}{(a+b+c+d)(a+b)}$$

Suponiendo que la prueba resultó positiva para ese paciente, entonces la probabilidad de enfermo dado que la prueba salió positiva se puede calcular con la fórmula de Bayes:

$$P(E|+) = \frac{P(E \cap +)P(E)}{P(+)} = \frac{\left(\frac{a}{a+b+c+d}\right)\left(\frac{a+c}{a+b+c+d}\right)}{\frac{a+b}{a+b+c+d}} = \frac{a(a+c)}{(a+b+c+d)(a+b)}$$

Esto equivale a suponer que el paciente que participó en la prueba fue seleccionado de manera aleatoria de la población de pacientes para que tenga sentido la $P(+)$.

De manera análoga se puede obtener $P(NE|-)$.

Comparación de intervalos de verosimilitud (paradigma en la instancia) con los intervalos de confianza (paradigma comportamental)

Se considera el caso cuando la normalidad de los estimadores no se cumple. Considere la estimación de proporciones pequeñas (θ) por verosimilitud. En una muestra de n elementos se tiene que x de ellos tienen cierta propiedad, *vgr*. enfermos de un padecimiento raro. Se supone una distribución binomial para x, el número de casos enfermos y θ es la proporción o prevalencia de la enfermedad en la población. θ es desconocida y se quiere conocer su posible valor. El modelo usado es:

$$P(X = x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

Para una x fija, este modelo se puede interpretar como la función que da la probabilidad de esa x para los posibles valores de θ. Esta distribución se supone en la mayoría de los casos dentro del paradigma comportamental.

El estimador de máxima verosimilitud es $\hat{\theta} = x/n$, por lo tanto, la $P(X=x)$ es máxima con ese valor como θ. Entonces, se puede calcular, para otros valores de

This is equivalent to assuming that the patient who participated in the test was selected randomly from the population of patients so that $P(+)$ will have sense.

Analogically, $P(NE|-)$ can be obtained.

Comparison of likelihood intervals (paradigm in the instance) versus confidence intervals (compartmental paradigm)

This case is considered when the normality of the estimators is not fulfilled. So, what should be done is to consider the estimator of small proportions (θ) by likelihood. For instance, in a sample of n elements, you have that x of them have certain property, *vgr.* patients with rare disease. If a binomial distribution for x is assumed, the number of cases that are sick, and θ is the proportion or prevalence of sickness in the population. If θ is unknown and it is desired to know the possible value, the model used is:

$$P(X = x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

For a fixed x, this model can be interpreted as the function that gives the probability of x for the possible values of θ. This distribution is assumed in most of the cases within the compartmental paradigm.

The estimator of maximum likelihood is $\hat{\theta} = x/n$, hence $P(X=x)$ is maximum with the value as θ. Thus, we can calculate how much this probability decreases for other values of θ. That is, $\hat{\theta}$ is the value of θ that most likely produces x. Any other value of θ can produce the sample with the same value of x but with less probability. But values of θ close to $\hat{\theta}$ produces a value of x with a probability a little less than that maximum value. So, we want to know those values of θ that "cannot detract", because they can produce x with good probability. They are "compatible" with the occurrence of x; that is, they are likely. To determine the values of θ compatibles with the data (the value of x), the standardized likelihood is constructed using the equation:

$$\frac{\theta^x(1-\theta)^{n-x}}{\hat{\theta}^x(1-\hat{\theta})^{n-x}} = \frac{L(\theta|x)}{L(\hat{\theta}|x)}$$

Figure 10 shows the standardized likelihood functions for binomial distributions with different

θ, qué tanto disminuye esa probabilidad. Es decir, $\hat{\theta}$ es el valor de θ que con más probabilidad produce x, y cualquier otro valor de θ produce la muestra con el mismo valor x con menor probabilidad. Pero valores de θ cercanos a $\hat{\theta}$ producen valor de x con probabilidad un poco menor que ese valor máximo. Así que se quiere conocer aquellos valores de θ que pueden "no desmerecer" porque pueden producir x con buena probabilidad. Son "compatibles" con la ocurrencia de x; es decir, son verosímiles. Para determinar valores de θ compatibles con los datos (el valor de x) se construye la verosimilitud estandarizada que está dada por:

$$\frac{\theta^x(1-\theta)^{n-x}}{\hat{\theta}^x(1-\hat{\theta})^{n-x}} = \frac{L(\theta|x)}{L(\hat{\theta}|x)}$$

La Figura 10 muestra las funciones de verosimilitud estandarizadas para distribuciones binomiales con diferentes valores para n y θ como parámetros. Cuando $x=2$ y $n=2$ de manera que el máximo valor de la verosimilitud ocurre cuando $\hat{\theta}=\frac{x}{n}=1$, se le llama estimador de máxima verosimilitud. Cuando $x=2$ y $n=3$, el estimador de máxima verosimilitud es: $\hat{\theta}=\frac{2}{3}=0.\overline{666}$

values for n and θ as parameters. When $x=2$ and $n=2$, then the maximum likelihood value would occur when $\hat{\theta}=\frac{x}{n}=1$, it is called maximum likelihood estimator. When $x=2$ and $n=3$, the maximum likelihood estimator is: $\hat{\theta}=\frac{2}{3}=0.\overline{666}$.

A case where confidence Interval of likelihood can be constructed but not that of the behavioral paradigm is when $x=0$, in this case $\hat{\theta}=0/n=0$. The likelihood approach does allow to obtain values of θ compatibles with the data $x=0$. It is assumed that the values of θ that have a likelihood equals or greater than 0.2 are compatible with the information. In Figure 11, three cases with $x=0$ and $x=5$, 10 and 20 are illustrated. Note that the amplitude of the likelihood interval reduces on increasing n.

Likelihood principle

Likelihood function is what gives evidence on the parameters, although the design is different. Thus, the evidence between a binomial and a negative binomial, for instance, is equal when the values of n and x are the same in the two designs, because the part that involves θ is the same in both expressions as can be seen below:

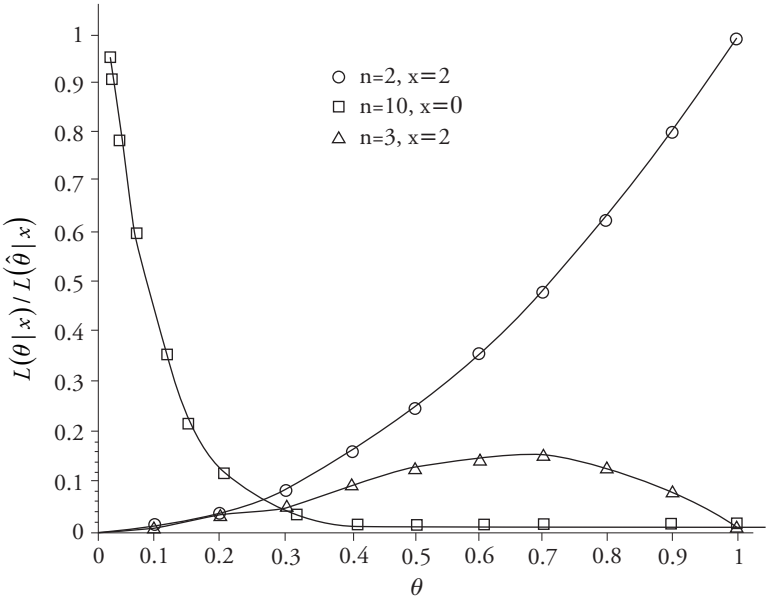

Figura 10. Funciones de verosimilitud para distribuciones binomiales.
Figure 10. Likelihood function for binomial distributions.

Un caso donde no se puede construir el intervalo de confianza del paradigma comportamental pero si el de verosimilitud, es cuando $x=0$, en este caso $\hat{\theta}=0/n=0$. El enfoque de verosimilitud sí permite obtener valores de θ compatibles con los datos $x=0$. Se tiene que valores de θ que tienen una verosimilitud igual o mayor que 0.2 son compatibles con la información. En la Figura 11 se ilustran tres casos, con $x=0$ y $n=5$, 10 y 20. Nótese que se reduce la amplitud del intervalo de verosimilitud al aumentar n.

Principio de verosimilitud

La función de verosimilitud es la que da evidencia sobre los parámetros, aunque el diseño sea diferente. Así, la evidencia entre una binomial y una binomial negativa, por ejemplo, es igual cuando los valores de n y x son los mismos en los dos diseños, porque la parte que involucra a q es la misma en ambas expresiones, como puede notarse a continuación:

$$P(X=x)=\binom{n}{x}\theta^x(1-\theta)^{n-x}$$

$$P(N=n)=\binom{n-1}{x-1}\theta^x(1-\theta)^{n-x}$$

$$P(X=x)=\binom{n}{x}\theta^x(1-\theta)^{n-x}$$

$$P(N=n)=\binom{n-1}{x-1}\theta^x(1-\theta)^{n-x}$$

CONCLUSIONS

In this document, we have presented some of the most commonly used probability and also the essential aspects of the three paradigms for statistical inference. It is emphasized that since paradigms are based on different assumption, they never compete with each other. Hence, it is possible to extrapolate ideas from one paradigm to the assumptions of another, if there is coherence between the specific assumptions of the paradigm that incorporates the ideas. Neyman and Pearson did it with the Fisher's likelihood concept. What is important is to know clearly which are the assumptions of each paradigm and do not forget that, regardless of the approach, the main objective of the inference is not the decision about the action in practice, but to gain knowledge about the true state of the things. Of course, if there is a great confidence that the state of things is true, decision can be made.

—End of English version—

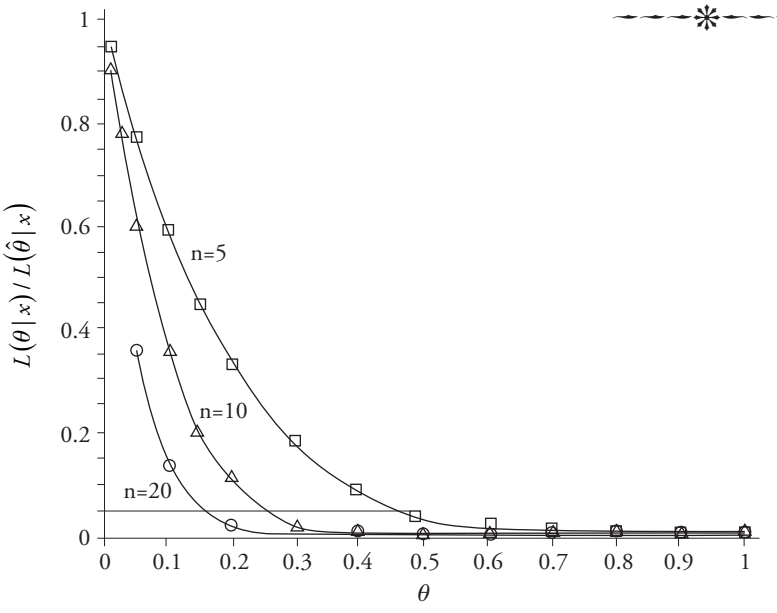

Figura 11. Intervalo de verosimilitud para una binomial con $x=0$ y $n=5$, 10 y 20.
Figure 11. Likelihood interval for a binomial with $x=0$ and $x=5$, 10 and 20.

CONCLUSIONES

En este documento hemos presentado algunos de los modelos de probabilidad más utilizados y también los aspectos esenciales de los tres paradigmas para la inferencia estadística. Se enfatiza que los paradigmas no compiten entre sí porque se basan en supuestos diferentes. Así, se pueden tomar ideas de uno y utilizarlas dentro de los supuestos de otro, siempre y cuando se tenga coherencia entre los supuestos específicos del paradigma que incorpora las ideas. Neyman y Pearson hicieron esto con el concepto de verosimilitud de Fisher. Lo importante es tener claro cuáles son los supuestos de cada paradigma y no olvidar que, independientemente del enfoque, el objetivo principal de la inferencia no es la decisión sobre una acción en la práctica, sino ganar conocimiento sobre el verdadero estado de las cosas. Por supuesto, si se considera con gran confianza que un estado de cosas es cierto, se pueden tomar decisiones.

LITERATURA CITADA

Chatterjee, S. K. 2003. Statistical Thought: A Perspective and History. Oxford University Press. 440 p.

Khun, T. S. 1996. La Estructura de las Revoluciones Científicas. Breviarios Fondo de Cultura Económica México. 264 p.

Méndez-Ramírez, I., H. Moreno-Macías, I. Méndez Gómez-Humarán, y C. Murata. 2013. Objetividad y Población. Monografía IIMAS UNAM. 54 p.

Pawitan, Y. 2001. In all Likelihood: Statistical Modeling and Inference using Likelihood. Oxford University Press. 544 p.

Popper, K. 1959. The Logic of Scientific Discovery, London: Hutchinson. 513 p.

Salsburg, D. 2001. The Lady Testing Tea. How Statistics Revolutionized Science in the Twentieth Century. W. H. Freeman and Company. New York. 352 p.

Thompson, B. 2007. The Nature of Statistical Evidence. Springer. 152 p.