ARTÍCULO ORIGINAL

Revista de Investigación de Física **27(3)**, (Set-Dic 2024)
Doi: 10.15381/rif.v27i3.23824

eISSN:1728-2977

Pruebas estadísticas en física de partículas

Mississippi Valenzuela *[1,2]

[1] *Universidad Nacional Autónoma de México, Instituto de Física, Departamento de Física Teórica, Apartado Postal 20-364, Cd. México 01000, México*
[2] *Universidad Autónoma de Zacatecas, Unidad Académica de Ciencia y Tecnología de la Luz y la Materia, Parque Científico y Tecnológico QUANTUM Ciudad del Conocimiento, Zacatecas 98160, México*

Recibido 17 Oct 2022 – Aceptado 14 Dic 2024 – Publicado 21 Dic 2024

Resumen
La física de partículas surgió como una disciplina pionera en la "gran ciencia"; se realizan experimentos en los aceleradores que aumentan cada vez más su energía y la complejidad de las colaboraciones de miles de físicos de cientos de institutos de unas decenas de países. En la "gran ciencia" se ha desarrollado una metodología de investigación en el que la estadística es de gran importancia para interpretar los resultados que se obtienen de estos experimentos. A pesar del descubrimiento del bosón de Higgs en el 2012, lo que contribuyó al éxito del Modelo Estándar y al otorgamiento del premio Nóbel de Física del 2013. Basándonos en estas motivaciones, en el presente trabajo se muestran ciertas pruebas estadísticas basadas en el método de "verosimilitud" y se aplican en la física de altas energías en el descubrimiento de nuevos fenómenos, al hacer uso de la construcción de intervalos de confianza de los parámetros del modelo. Nos centramos en las propiedades de los métodos llamados "pruebas" estadísticas que permiten incluir las incertidumbres sistemáticas. También, trabajamos con fórmulas explícitas para las distribuciones asintóticas de las pruebas estadísticas. Además, usaremos un conjunto de datos representativo, llamado el conjunto de datos "Asimov", que proporciona un método para obtener la sensibilidad experimental de una búsqueda o medición, así como las fluctuaciones en esa espera.
Palabras clave: Prueba estadística, función de verosimilitud, función de distribución de probabilidad.

Statistical tests in particle physics

Abstract
Particle physics emerged as a pioneer discipline in "big science"; experiments are being carried out in accelerators that are increasing in energy and complexity by the collaborations of thousands of physicists from hundreds of institutes in a few dozen countries. In "big science" a research methodology has been developed in which statistics is of great importance to interpret the results obtained from these experiments. Despite the discovery of the Higgs boson in 2012, which contributed to the success of the Standard Model and the award of the 2013 Nobel Prize in Physics. Based on these motivations, in the present work certain statistical tests based on the "likelihood" method are shown and are applied in high energy physics in the discovery of new phenomena, by making use of the construction of confidence intervals of the model parameters. We focus on properties of methods called statistical "proofs" that allow systematic uncertainties to be included. Also, we work with explicit formulas for the asymptotic distributions of the statistical tests. In addition, we will use a representative data set, called the "Asimov" data set, which provides a method for obtaining the experimental sensitivity of a search or measurement, as well as the fluctuations in that expectation.
Keywords: Statistical test, likelihood function, probability distribution function.

*mvalenzuelalumat@uaz.edu.mx

1. Introducción

El objetivo principal de análisis de datos en Física de Altas Energías es comprender las interacciones de las partículas a través de experimentos y, al hacerlo, se buscan fenómenos que van más allá del Modelo Estándar. En experimentos de física de partículas se buscan los procesos predichos por los teóricos de la física que aún no han sido vistos. En dichos experimentos, la significación estadística de una señal observada se cuantifica por medio del valor p.

En este trabajo se investiga el método para obtener fórmulas explícitas de distribuciones asintóticas. Dichas distribuciones se basan en los resultados debido a Wilks [1] y Wald [2]. Mediante los trabajos de Wilks y Wald se obtiene la significación estadística de los datos proporcionados, así como la distribución completa del muestreo de la significación mediante las hipótesis de los diferentes modelos de señal, todo sin recurrir al método de Monte Carlo [3, 4].

Un elemento útil de estos métodos implica la estimación de la mediana reemplazando el conjunto de datos simulados por un sólo representante, es decir, el conjunto de datos "Asimov". En años pasados, estos métodos se han utilizado y justificado intuitivamente [5, 6]. En el presente trabajo se proporciona una justificación matemática para el método [6]. Así, también se exploran sus limitaciones y se señalan varios aspectos adicionales de su uso. Además, se extiende el trabajo que se muestra en la referencia [7]. La extensión se hace al mostrar fórmulas más precisas para la importancia de exclusión (rechazo) y proporcionar una medida cuantitativa de las fluctuaciones estadísticas en la significación de descubrimiento y límites de rechazo.

2. Pruebas de hipótesis

Una de las tareas fundamentales del análisis estadístico es probar si la predicciones de un determinado modelo están de acuerdo con los datos observados. Aquí se utilizará $\vec{x}$ para indicar el resultado de una medición; podría representar una sola cantidad o una colección de valores. Una hipótesis H significa una afirmación para la probabilidad de encontrar los datos $\vec{x}$ (o si $\vec{x}$ incluye las variables continuas, H especifica una función de densidad de probabilidad o pdf). Además, se escribirá $P\left(\vec{x}|H\right)$ para la probabilidad de encontrar los datos $\vec{x}$ ante la suposición de la hipótesis H.

Consideremos una hipótesis H_0 que se pone a prueba (vamos a menudo llamar a esto la hipótesis "nula") y una hipótesis alternativa H_1. En la estadística frecuentista se define una prueba de H_0 al especificar un subconjunto del espacio de datos denominada la "región crítica", w, de manera que la probabilidad de observar los datos allí satisface la desigualdad

$$P\left(\vec{x} \in w|H_0\right) \leq \alpha. \tag{1}$$

En la expresión (1) α es una constante especifica antes de llevar a cabo la prueba, normalmente se establece por convención para un valor pequeño de 5 %. Para los datos continuos, se toma la relación (1) como una igualdad. Si los datos son discretos, tales como una serie de eventos, entonces es posible que no exista ningún subconjunto de los valores de los datos cuya probabilidad sumada es exactamente igual a alfa, por lo que se lleva la región crítica para tener una probabilidad hasta α. La región crítica w define la prueba. Si se observan los datos en w entonces se rechaza la hipótesis H_0.

Hasta este momento, la ecuación (1) es la única propiedad que define la prueba. Dicha expresión establece que la probabilidad de encontrar los datos en la región crítica no es más que α. Pero hay muchos, si no un número infinito de posibles subconjuntos del espacio de datos que satisfacen este criterio, y no está claro que debería tomarse como la región crítica. Aquí es donde la hipótesis alternativa H_1 entra en juego (ver figura 1).

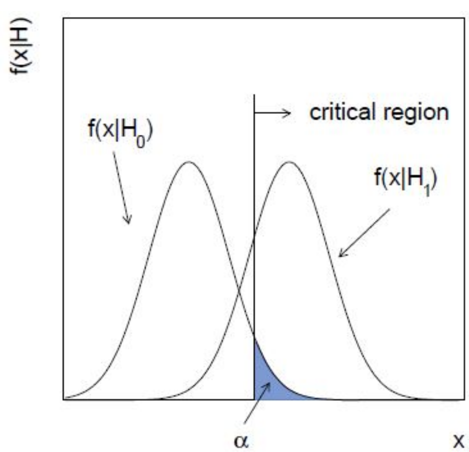

Figura 1: Ilustración de la región crítica de una prueba estadística.

Al rechazar la hipótesis H_0, se tiene un error de tipo I. Mediante la construción de la probabilidad de que esto ocurra es el tamaño de la prueba, α. Si, por otro lado, no rechazamos H_0, debido a que la hipótesis H_1 resultó ser cierta, entonces se tiene un error tipo II. La probabilidad de rechazar la hipótesis nula si la hipótesis alternativa H_1 es verdadera se llama la "potencia" de la prueba con respecto a H_1, que es la unidad menos la probabilidad de un error tipo II.

Una "prueba de significación" de una hipótesis H está estrechamente relacionada con las pruebas descritas anteriormente. Se supone una medición de resultados en los datos $\overrightarrow{\mathbf{x}}$ (un número único o una colección de muchos valores) para el cual la hipótesis H predice la probabilidad $P\left(\overrightarrow{\mathbf{x}}|H\right)$.

El analista especifica que valores posibles de los datos constituirían un nivel de incompatibilidad con H que es igual o mayor que la existente entre H y los datos observados $\overrightarrow{\mathbf{x}}_{obs}$. Una vez que esto se da, entonces se calcula el valor p de H como la probabilidad para encontrar datos en esta región de igual o mayor incompatibilidad.

Cuando se dice que un dado $\overrightarrow{\mathbf{x}}$ tiene menos compatibilidad con H, se implica que tiene una mayor compatibilidad con alguna hipótesis alternativa. Esto es análogo a la ambigüedad encontrada en la determinación de la región crítica de una prueba.

En una prueba frecuentista, se rechaza H_0 si los datos se encuentran en la región crítica, o lo que es equivalente, si el valor p de H_0 es menor o igual a α. A pesar de este lenguaje, no es necesariamente cierto creer que H_0 es falsa. Para hacer esta afirmación se cuantifica nuestro nivel de confianza acerca de H_0 para la probabilidad subjetiva, y se calcula mediante el teorema de Bayes, es decir,

$$P\left(H_0|\overrightarrow{\mathbf{x}}\right)=\frac{P\left(\overrightarrow{\mathbf{x}}|H_0\right)\pi\left(H_0\right)}{\sum_i P\left(\overrightarrow{\mathbf{x}}|H_i\right)\pi\left(H_i\right)}. \tag{2}$$

La probabilidad a posteriori $P\left(H_0|\overrightarrow{\mathbf{x}}\right)$ es proporcional a la probabilidad a priori $\pi\left(H_0\right)$, y esto necesaríamente debe ser especificado si queremos expresar nuestro nivel de confianza de que la hipótesis es verdadera.

También es importante señalar que el valor p de una hipótesis H_0 no es el mismo como la probabilidad (2) que es verdad, sino la probabilidad, bajo el supuesto de H_0, para encontrar datos con al menos tanta incompatibilidad con H_0 como los datos que se encontraron realmente. El valor p, por lo tanto, no depende de las probabilidades a priori.

3. Formalismo de una búsqueda como una prueba estadística

En esta sección se describe el procedimiento general que se utiliza para buscar un nuevo fenómeno en el contexto de una prueba estadística frecuentista. A los efectos de descubrir un nuevo proceso de señal, se define la hipótesis nula, H_0, como la descripción sólo para procesos conocidos, aquí es designada como fondo. Esto se va a probar en contra de la hipótesis alterna H_1, que incluye tanto el fondo como la señal. Al establecer límites, el modelo con la señal más el fondo desempeña el papel de H_0, que se prueba contra la hipótesis de fondo, H_1.

En la física de partículas usualmente se convierte el valor p en la significancia equivalente, Z. Z se define de tal manera que una variable distribuida en forma Gaussiana se encuentra a Z desviaciones estándar[1] a la derecha de su media que tiene una probabilidad (en la cola superior) igual a p. Esto es,

$$Z=\Phi^{-1}\left(1-p\right), \tag{3}$$

donde Φ^{-1} es el cuántil (inversa de la distribución acumulativa) de la distribución Gaussiana estándar. Para un proceso de señal, como el bosón de Higgs, la comunidad de la física de partículas ha considerado el rechazo de la hipótesis de fondo con una significancia de, al menos, $Z=5$ como un nivel apropiado para constituir un descubrimiento. Esto corresponde a $p=2.87\times10^{-7}$. Para propósito de la exclusión de una hipótesis de la señal, un valor límite de p de 0.05 (es decir, 95 % de nivel de confianza) corresponde a $Z=1.64$.

Debe hacerse hincapié en que, en un contexto científico actual, el rechazo de la hipótesis de fondo sólo en un sentido estadístico es sólo una parte del descubrimiento de un nuevo fenómeno. Aquí, sin embargo, sólo tenemos en cuenta la tarea de determinar el valor p de la hipótesis de fondo; si se encuentra por debajo de un umbral determinado, entonces se considera esto como "descubrimiento".

A menudo es útil cuantificar la sensibilidad de un experimento mediante la información de la significación esperada[2]. Dicha sensibilidad se obtendrá con una medición dada bajo el supuesto de varias hipótesis. Por ejemplo, la sensibilidad para el descubrimiento de un proceso de señal H_1 se caracteriza a través del valor esperado, bajo el supuesto de H_1, del valor de Z obtenido de una prueba de la hipótesis H_0.

Un procedimiento ampliamente utilizado para establecer el descubrimiento (o exclusión) en física de partículas se basa en una prueba de significación frecuentista usando el cociente de verosimilitud como una prueba estadística. Además de los parámetros de interés, tales como la tasa (sección transversal) del proceso de la señal, los modelos de señal y de fondo contendrán parámetros de perturbación generales cuyos valores no son conocidos a priori, sino que serán ajustados de los datos.

Para ilustrar el uso del cociente de verosimilitud, se considera un experimento en el que para cada evento seleccionado se miden los valores de ciertas variables cinemáticas, y por lo tanto los datos resultantes pueden ser representados como uno o más histogramas. Además, se supone que para cada evento en la muestra de la señal se mide una variable x y se usan estos valores para construir un histograma $\overrightarrow{\mathbf{n}}=(n_1,n_2,...,n_N)$. El valor esperado de

[1] Algunos autores han definido; por ejemplo, Ref. [8], ésta relación usando una fluctuación en las dos colas de una variable Gaussiana con una significancia de 5σ para un valor de $p=5.7\times10^{-7}$. Se toma la definición de uno de los lados de la variable Gaussiana ya que $Z=0$ y $p=0.5$.

[2] Durante el resto del artículo, el término "significancia esperada" se usará siempre para referirnos a la mediana.

n_i puede ser escrito mediante

$$E[n_i] = \mu s_i + b_i, \tag{4}$$

donde el número medio de entradas en el i-ésimo intervalo de la señal y el fondo son

$$s_i = s_{total} \int_i f_s\left(x; \overrightarrow{\theta}_s\right) dx, \tag{5}$$

$$b_i = b_{total} \int_i f_b\left(x; \overrightarrow{\theta}_b\right) dx. \tag{6}$$

Aquí el parámetro μ determina la intensidad de la señal del proceso, con $\mu = 0$ corresponde sólo a la hipótesis de fondo. Para $\mu = 1$ se tiene la hipótesis de la señal nominal. Las funciones $f_s\left(x; \overrightarrow{\theta}_s\right)$ y $f_b\left(x; \overrightarrow{\theta}_b\right)$ son las funciones de densidad de probabilidad (pdfs) de la variable x para eventos de la señal y del fondo, y $\overrightarrow{\theta}_s$ y $\overrightarrow{\theta}_b$ representan los parámetros que caracterizan las formas de las pdfs. Las cantidades s_{total} y b_{total} son los números medios totales de la señal y del fondo, y las integrales en (5) y (6) representan las probabilidades de un evento que se encuentran en un intervalo i. A continuación se utilizará $\overrightarrow{\theta} = \left(\overrightarrow{\theta}_s, \overrightarrow{\theta}_b, b_{total}\right)$ para denotar todos los parámetros de perturbación. La señal de normalización s_{total} no es, sin embargo, un parámetro ajustable, sino más bien se fija al valor predicho por el modelo de señal nominal.

Además del histograma medido $\overrightarrow{\mathbf{n}} = (n_1, n_2, ..., n_N)$ a menudo se hacen más mediciones auxiliares que ayudan a limitar los parámetros de perturbación. Por ejemplo, se selecciona una muestra de control, donde se esperan principalmente eventos de fondo y de ellos construir un histograma de alguna variable cinemática elegida. Esto da a continuación un conjunto de valores $\overrightarrow{\mathbf{m}} = (m_1, m_2, ..., m_M)$ para el número de entradas en uno de los M intervalos. El valor esperado de m_i puede ser escrito como

$$E[m_i] = u_i\left(\overrightarrow{\theta}\right), \tag{7}$$

donde las u_i son cantidades calculables en función de los parámetros $\overrightarrow{\theta}$. A menudo se construye esta medición a fin de proporcionar información sobre el parámetro b_{total} y también posiblemente de los parámetros de la señal y de fondo.

La función de verosimilitud es el producto de las funciones de probabilidades de Poisson para todos los intervalos [8, 9]:

$$\mathcal{L}\left(\mu, \overrightarrow{\theta}\right) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_{k=1}^{M} \frac{u^{m_k}}{m_k!} e^{-u_k}. \tag{8}$$

Para probar la hipótesis de un valor de μ se tiene en cuenta la razón de verosimilitud [9, 10]

$$\lambda(\mu) = \frac{\mathcal{L}\left(\mu, \widehat{\widehat{\overrightarrow{\theta}}}\right)}{\mathcal{L}\left(\widehat{\mu}, \widehat{\overrightarrow{\theta}}\right)}. \tag{9}$$

El numerador de esta relación es la función de verosimilitud perfil [11]. Aquí $\widehat{\widehat{\overrightarrow{\theta}}}$ en el numerador denota el valor de $\overrightarrow{\theta}$ que maximiza a $\mathcal{L}$ para la μ especificada, es decir, es el estimador condicional de máxima verosimilitud (ML) de $\overrightarrow{\theta}$ (y por lo tanto es una función de μ). El denominador es la función de verosimilitud maximizada, es decir, $\widehat{\mu}$ y $\widehat{\overrightarrow{\theta}}$ son sus estimadores ML. Esto refleja la pérdida de información sobre μ debido a las incertidumbres sistemáticas.

En muchos análisis, la contribución del proceso de señal al número medio de eventos se supone que es no negativo. Esta condición implica efectivamente que cualquier estimador físico para μ debe ser no negativo. Incluso si consideramos que este sea el caso, sin embargo, es conveniente definir un estimador eficaz $\widehat{\mu}$ como el valor de μ que maximiza la verosimilitud, es decir, esto da $\widehat{\mu} < 0$ (pero siempre que los valores medios Poissonianos, $\mu s_i + b_i$, sean no negativos). Dichas condiciones se permitirán en la subsección 4.1 modelar a $\widehat{\mu}$ como una distribución Gaussiana, y de esta manera se determina la distribución de las pruebas estadísticas que consideramos.

3.1. Prueba estadística $t_\mu = -2\ln\lambda(\mu)$

De la ecuación (9) se ve que $0 \leq \lambda \leq 1$, con λ cerca de 1, lo que implica una buena concordancia entre los datos y el valor de la hipótesis de μ. De manera equivalentemente se usa el estadístico

$$t_\mu = -2\ln\lambda(\mu) \tag{10}$$

como la base de una prueba estadística. Los valores más altos de t_μ corresponden al aumento de la incompatibilidad entre los datos y μ.

Una prueba de un valor hipotético de μ se define mediante el uso del estadístico t_μ directamente como medida de la discrepancia entre los datos y la hipótesis, con valores más altos de t_μ correspondientes al incremento de desacuerdo. Para cuantificar el nivel de desacuerdo se calcula el valor de p,

$$p_\mu = \int_{t_{\mu,obs}}^{\infty} f(t_\mu|\mu)\, dt_\mu, \tag{11}$$

donde $t_{\mu,obs}$ es el valor del estadístico t_μ observado de los datos y $f(t_\mu|\mu)$ denota la pdf de t_μ bajo el supuesto de la intensidad de la señal μ. Las aproximaciones útiles para este y otras pdfs relacionadas se dan en la sección 4.

Cuando se utiliza el estadístico t_μ, un conjunto de datos puede resultar en un valor p de dos maneras distintas: la intensidad de la señal estimada $\widehat{\mu}$ puede ser mayor o menor que el valor planteado de la hipótesis de μ. Como resultado, el conjunto de los valores μ que se rechazó debido a que sus valores de p se encuentran por debajo de un umbral especificado α puede estar a cualquier lado de esos valores no rechazados, i. e., se puede obtener un intervalo de confianza de dos lados (colas) para μ.

3.2. Prueba estadística $\widetilde{t}_\mu = -2\ln\widetilde{\lambda}(\mu)$ para $\mu \geq 0$

La presencia de una nueva señal sólo puede aumentar la tasa media de eventos más allá de la que se espera únicamente del fondo. Es decir, el proceso de señal tiene necesariamente $\mu \geq 0$, y se define una prueba estadística alternativa denominada $\widetilde{t}_\mu$.

Para un modelo donde $\mu \geq 0$, si se encuentran los datos de tal manera que $\mu < 0$, entonces el mejor nivel de acuerdo entre los datos y cualquier valor físico de μ ocurre para $\mu = 0$. Por consiguiente, se tiene la expresión siguiente

$$\widetilde{\lambda}(\mu) = \begin{cases} \dfrac{\mathcal{L}\left(\mu, \widehat{\widehat{\overrightarrow{\theta}}}(\mu)\right)}{\mathcal{L}\left(\widehat{\mu}, \widehat{\overrightarrow{\theta}}\right)}, & \widehat{\mu} \geq 0, \\ \dfrac{\mathcal{L}\left(\mu, \widehat{\widehat{\overrightarrow{\theta}}}(\mu)\right)}{\mathcal{L}\left(0, \widehat{\widehat{\overrightarrow{\theta}}}(0)\right)}, & \widehat{\mu} < 0. \end{cases} \tag{12}$$

Aquí $\widehat{\widehat{\overrightarrow{\theta}}}(0)$ y $\widehat{\widehat{\overrightarrow{\theta}}}(\mu)$ se refieren a los estimadores de máxima verosimilitud de $\overrightarrow{\theta}$ dado un parámetro de intensidad de 0 o μ, respectivamente.

En la ecuación (10) la variable $\widetilde{\lambda}(\mu)$ se utiliza en lugar de $\lambda(\mu)$. Así, se obtiene la correspondiente prueba estadística $\widetilde{t}_\mu$. Esto es, se escribe la expresión

$$\widetilde{t}_\mu = -2\ln\widetilde{\lambda}(\mu) = \begin{cases} -2\ln\dfrac{\mathcal{L}\left(\mu, \widehat{\widehat{\overrightarrow{\theta}}}(\mu)\right)}{\mathcal{L}\left(\widehat{\mu}, \widehat{\overrightarrow{\theta}}\right)}, & \widehat{\mu} \geq 0, \\ -2\ln\dfrac{\mathcal{L}\left(\mu, \widehat{\widehat{\overrightarrow{\theta}}}(\mu)\right)}{\mathcal{L}\left(0, \widehat{\widehat{\overrightarrow{\theta}}}(0)\right)}, & \widehat{\mu} < 0. \end{cases} \tag{13}$$

Como se hizo con la estadística t_μ, se cuantifica el nivel de desacuerdo entre los datos y el valor hipotético μ con el valor p, al igual que en la ecuación (11). Para esto se necesita la distribución de t_μ, una aproximación de dicha distribución se proporciona en la subsección 4.4.

De manera similar al caso de t_μ, los valores de μ por encima y debajo de $\widehat{\mu}$ pueden ser excluidos por un determinado conjunto de datos, es decir, se puede obtener un intervalo de confianza unilateral, o bien, uno de dos colas para μ. En el caso de que no haya parámetros de perturbación, la variable de contraste $\widetilde{t}_\mu$ es equivalente a la utilizada en la construcción de intervalos de confianza acorde al procedimiento de Feldman y Cousins [12].

3.3. Prueba estadística q_0

Un caso importante de la prueba estadística t_μ se utiliza para la prueba $\mu = 0$, es decir, en un modelo donde se asume $\mu \geq 0$. Al rechazar la hipótesis $\mu = 0$, se conduce al descubrimiento de una nueva señal, respectivamente. En este importante caso se usa la notación especial $q_0 = t_0$. Al utilizar la definición (13) con $\mu = 0$ se encuentra

$$q_0 = \begin{cases} -2\ln\lambda(0), & \mu \geq 0, \\ 0, & \mu < 0, \end{cases} \tag{14}$$

donde $\lambda(0)$ es la razón de verosimilitud para $\mu = 0$ como se ha definido en la ecuación (9).

La ecuación (14) se puede contrastar con t_0, es decir, la ecuación (10), dicha se uso para probar $\mu = 0$. En este caso, se puede rechazar la hipótesis $\mu = 0$, ya sea para una fluctuación hacia arriba o hacia abajo de los datos. Esto es adecuado si la presencia de un nuevo fenómeno podría conducir a un aumento o disminución en el número de eventos que se encuentran.

Al utilizar q_0, se tienen en cuenta los datos para mostrar la falta de acuerdo con la hipótesis de fondo sólo si $\widehat{\mu} > 0$. Es decir, un valor de $\widehat{\mu}$ muy por debajo de cero, de hecho puede constituir evidencia en contra del modelo de fondo, pero este tipo de discrepancia no muestra que los datos contienen eventos de señal, sino que apunta a algún otro error sistemático. Para la presente discusión se supone que las incertidumbres sistemáticas son tratadas por los parámetros de perturbación $\overrightarrow{\theta}$.

Para cuantificar el nivel de desacuerdo entre los datos y la hipótesis de $\mu = 0$ usando el valor observado de q_0 calculamos el valor p de la misma manera como lo hacimos con t_μ, es decir,

$$p_0 = \int_{q_{0,obs}}^{\infty} f(q_0|0)\, dq_0. \tag{15}$$

Aquí $f(q_0|0)$ denota la función de densidad de probabilidad de la prueba estadística q_0 ante la suposición de la hipótesis de fondo ($\mu = 0$). Una aproximación para función y otras pdfs relacionadas se encuentran en la subsección 4.5.

3.4. Prueba estadística q_μ para límites superiores

Al establecer un límite superior para el parámetro μ, se consideran dos pruebas estadísticas estrechamente relacionadas. En primer lugar, se define

$$q_\mu = \begin{cases} -2\ln\lambda(\mu), & \widehat{\mu} \leq \mu, \\ 0, & \widehat{\mu} > \mu, \end{cases} \tag{16}$$

donde $\lambda(\mu)$ es la razón de verosimilitud como se muestra en la ecuación (9). La razón para establecer $q_\mu = 0$ en el caso $\widehat{\mu} > \mu$ se debe al establecer un límite superior, donde no se considerarían datos con $\widehat{\mu} > \mu$, y por lo tanto esto no se toma como parte de la región de rechazo de la prueba. En virtud de la prueba estadística (16) los valores más altos de q_μ representan mayor incompatibilidad entre los datos y el valor hipotético de μ.

q_0 no es simplemente un caso especial de $q_\mu$con $\mu = 0$, sino que tiene una definición diferente (ver las ecuaciones (14) y (16)). Es decir, q_0 es cero si los datos fluctúan hacia abajo ($\widehat{\mu} < 0$), pero q_μ es cero si los datos fluctúan hacia arriba ($\widehat{\mu} > \mu$). Con esta advertencia en mente, en el resto, q_μ se identificará con q_0 o q_μ según sea el contexto.

En esta prueba estadística se cuantifica el grado de concordancia entre los datos y μ con el valor p. Para un valor de q_μ observado se tiene

$$p_\mu = \int_{q_{\mu,obs}}^{\infty} f\left(q_\mu, \mu\right) dq_\mu, \tag{17}$$

el cual se expresa como una significancia al usar la ecuación (3). $f\left(q_\mu|\mu\right)$ es la función de densidad de probabilidad de q_μ al asumir la hipótesis μ. En la subsección 4.6 se proporcionan aproximaciones útiles para esta prueba estadística y otras pdf relacionadas.

3.5. Prueba estadística $\widetilde{q}_\mu$ alternativa para límites superiores

Si se consideran modelos para lo cuales $\mu \geq 0$, entonces la variable $\widetilde{\lambda}(\mu)$ se usa en lugar de $\lambda(\mu)$ en la ecuación (16) para obtener la estadística de prueba correspondiente, que denotamos por q_μ. Es decir,

$$\widetilde{q}_\mu = \begin{cases} -2\ln\frac{\mathcal{L}\left(\mu, \widehat{\widehat{\overrightarrow{\theta}}}(\mu)\right)}{\mathcal{L}\left(0, \widehat{\widehat{\overrightarrow{\theta}}}(0)\right)}, & \widehat{\mu} < 0, \\ -2\ln\frac{\mathcal{L}\left(\mu, \widehat{\widehat{\overrightarrow{\theta}}}(\mu)\right)}{\mathcal{L}\left(\widehat{\mu}, \widehat{\overrightarrow{\theta}}\right)} & 0 \leq \widehat{\mu} \leq \mu, \\ 0 & \widehat{\mu} > \mu. \end{cases} \tag{18}$$

Damos una aproximación para la pdf $f\left(\widetilde{q}_\mu|\mu'\right)$ en la subsección 4.7.

En los ejemplos numéricos [9] la diferencia entre las pruebas basadas en q_μ (Ecuación (16)) y $\widetilde{q}_\mu$ usualmente es insignificante, pero el uso de q_μ conduce a importantes simplificaciones.

4. Distribuciones asintóticas muestrales

Al encontrar el valor p de una hipótesis se usan las ecuaciones (15) o (17). Para dicho valor p se requiere la distribución de muestreo para la prueba estadística adecuada. En el caso del descubrimiento se prueba la hipótesis de fondo ($\mu = 0$), y por lo tanto se tiene que $f\left(q_0|0\right)$, donde q_0 se define por la ecuación (14). Al probar un valor distinto de cero de μ, para encontrar un límite superior se necesita la distribución $f\left(q_\mu|\mu\right)$, donde q_μ se define por la ecuación (16), o, también, se requiere la pdf de la prueba estadística $\widetilde{q}_\mu$ correspondiente a la ecuación (18). El subíndice de q se refiere a la hipótesis de prueba, y el segundo argumento de $f\left(q_\mu|\mu\right)$ proporciona el valor de μ asumido en la distribución de los datos.

También se necesita la distribución $f\left(q_\mu|\mu'\right)$ con $\mu \neq \mu'$ para encontrar qué significancia se puede esperar y cómo se comporta la distribución si los datos corresponden a un parámetro diferente del que se desea probar. Por ejemplo, al caracterizar la sensibilidad de un experimento planeado se cita la significancia de la mediana. Así se asumen los datos distribuidos de acuerdo con un modelo de señal especificado, con el cual se rechaza la hipótesis de fondo. Para esto se necesitaría $f\left(q_0|\mu'\right)$, con $\mu' = 1$. A partir de esto se puede encontrar la mediana de q_0, y por lo tanto la significancia del descubrimiento de la mediana. Al considerar la prueba estadística para límites superiores, se cita el valor de μ para el valor p de la mediana igual a 0.05, ya que esto da el límite superior de la mediana de μ con un nivel de confianza del 95 %. En este caso se necesita $f\left(q_\mu|0\right)$ (o, alternativamente, $f\left(\widetilde{q}_\mu|0\right)$).

4.1. Distribución aproximada de la razón de verosimilitud

En esta subsección se considera una prueba con el parámetro μ, el cual es o bien cero (para el descubrimiento) o distinto de cero (por un límite superior), y se supone que los datos se distribuyen de acuerdo con un parámetro μ'. La distribución deseada $f\left(q_\mu|\mu'\right)$ se encuentra al utilizar un resultado debido a Wald [2]. Wald mostró que para un único parámetro de interés,

$$-2\ln\lambda(\mu) = \frac{(\mu - \widehat{\mu})^2}{\sigma^2} + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right). \tag{19}$$

$\widehat{\mu}$ sigue una distribución Gaussiana con una media μ' y la desviación estándar σ, y N representa el tamaño de la muestra de datos. La desviación estándar σ de $\widehat{\mu}$ se obtiene de la matriz de covarianza de los estimadores para todos los parámetros, $V_{ij} = \mathrm{cov}\left[\widehat{\theta}_i, \widehat{\theta}_j\right]$, donde los θ_i representan tanto μ como los parámetros de perturbación (por ejemplo, tomar $\theta_0 = \mu$, así $\sigma^2 = V_{00}$). En el límite de la muestras grandes, el sesgo de los estimadores de máxima verosimilitud (ML) tiende a cero, en cuyo caso se escribe la inversa de la matriz de covarianza de la forma siguiente

$$V_{ij}^{-1} = -E\left[\frac{\partial^2 \ln \mathcal{L}}{\partial\theta_i \partial\theta_j}\right], \tag{20}$$

donde el valor esperado asume un parámetro μ'. Las aproximaciones que se presentan aquí son válidas cuando se desprecia el término $\mathcal{O}\left(1/\sqrt{N}\right)$, y el valor de σ se estima al usar la ecuación (20).

Si $\widehat{\mu}$ se distribuye de forma Gaussiana y despreciamos el término $\mathcal{O}\left(1/\sqrt{N}\right)$ de la ecuación (19), entonces la prueba estadística $t_\mu = -2\ln\lambda\left(\mu\right)$ sigue una distribución chi-cuadrada no centrada con un grado de libertad (véase, por ejemplo, [13]),

$$f\left(t_\mu;\Lambda\right) = \frac{1}{2\sqrt{t_\mu}}\frac{1}{\sqrt{2\pi}}\left\{\exp\left[-\frac{1}{2}\left(\sqrt{t_\mu}+\sqrt{\Lambda}\right)^2\right] + \exp\left[-\frac{1}{2}\left(\sqrt{t_\mu}-\sqrt{\Lambda}\right)^2\right]\right\},$$

donde el parámetro no central Λ es

$$\Lambda = \frac{\left(\mu-\mu'\right)^2}{\sigma^2}. \tag{21}$$

Para el caso especial $\mu' = \mu$ se tiene que $\Lambda = 0$ y $-2\ln\lambda\left(\mu\right)$ se aproxima a una distribución chi-cuadrada con un grado de libertad, un resultado mostrado por Wilks [1].

Los resultados de Wilks y Wald generalizan a más de un parámetro de interés. Si los parámetros de interés se pueden identificar de forma explícita con un subconjunto de los parámetros $\overrightarrow{\theta}_r = \left(\theta_1, ..., \theta_r\right)$, entonces la distribución $-2\ln\lambda\left(\overrightarrow{\theta}_r\right)$ sigue una distribución chi-cuadrada no central con r-grados de la libertad

$$\Lambda_r = \sum_{i=1}^{r}\sum_{j=1}^{r}\left(\theta_i - \theta_i'\right)\widetilde{V}_{ij}^{-1}\left(\theta_j - \theta_j'\right), \tag{22}$$

donde $\widetilde{V}_{ij}^{-1}$ es la inversa de la submatriz que se obtiene a partir de la restricción de la matriz de covarianza completa para los parámetros de interés. La matriz de covarianza completa se da a partir de la inversión de la ecuación (20), y se muestra una forma eficiente para calcularla en la subsección 4.2.

4.2. El conjunto de datos Asimov y la varianza de $\widehat{\mu}$

Algunas de las fórmulas requieren la desviación estándar σ de $\widehat{\mu}$, la cual se asume para una distribución Gaussiana con una media μ'. A continuación se muestran dos formas de estimar σ, las cuales están estrechamente relacionados con un conjunto de datos especial (artificial) denominado el conjunto de datos “Asimov".

Si se define el conjunto de datos Asimov para evaluar los estimadores de todos los parámetros, entonces se obtienen los verdaderos valores de los parámetros. La función de verosimilitud se considera en el análisis de la ecuación (8). La notación en esta sección se simplifica si se considera la ecuación siguiente

$$\nu_i = \mu' s_i + b_i. \tag{23}$$

Además $\theta_0 = \mu$ representa el parámetro de intensidad, de modo que $\overrightarrow{\theta}_i$ se acepta para cualquiera de los parámetros. Los estimadores ML para los parámetros se encuentran mediante las derivadas de $\ln\mathcal{L}$ con respecto a todos los parámetros igual a cero:

$$\frac{\partial \ln\mathcal{L}}{\partial\theta_j} = \sum_{i=1}^{N}\left(\frac{n_i}{\nu_i}-1\right)\frac{\partial\nu_i}{\partial\theta_j} + \sum_{i=1}^{M}\left(\frac{m_i}{u_i}-1\right)\frac{\partial u_i}{\partial\theta_j} = 0. \tag{24}$$

La ecuación (25) se cumple si los datos Asimov ($n_{i,A}$ y $m_{i,A}$) son iguales a sus valores esperados:

$$n_{i,A} = E\left[n_i\right] = \nu_i = \mu' s_i\left(\overrightarrow{\theta}\right) + b_i\left(\overrightarrow{\theta}\right) \tag{25}$$

$$m_{i,A} = E\left[m_i\right] = u_i\left(\overrightarrow{\theta}\right). \tag{26}$$

Aquí los valores paramétricos representan aquellos implicados por la distribución de los datos. En la práctica, estos son los valores que se estiman a partir del modelo de Monte Carlo utilizando una muestra de datos muy grande.

El conjunto de datos Asimov establecido se utiliza para evaluar la "verosimilitud Asimov" $\mathcal{L}_A$ y la correspondiente razón de verosimilitud λ_A. Los valores no enteros de los datos no es un problema, ya que los términos factoriales en la verosimilitud de Poisson representan constantes que se cancelan cuando se forma el cociente de verosimilitud, y por lo tanto sería más pequeño. En consecuencia, se encuentra que

$$\lambda_A = \frac{\mathcal{L}_A\left(\mu, \widehat{\widehat{\overrightarrow{\theta}}}\right)}{\mathcal{L}_A\left(\widehat{\mu}, \widehat{\overrightarrow{\theta}}\right)} = \frac{\mathcal{L}_A\left(\mu, \widehat{\widehat{\overrightarrow{\theta}}}\right)}{\mathcal{L}_A\left(\mu', \overrightarrow{\theta}\right)}, \tag{27}$$

donde la igualdad final usa el hecho de que los estimadores de los parámetros son iguales a sus valores hipotéticos cuando la verosimilitud se evalúa con el conjunto de datos Asimov.

Para encontrar σ se estima la matriz de la segunda derivada del logaritmo $\mathcal{L}$ (véase la ecuación (20)) para obtener la matriz inversa de la covarianza V^{-1}, invirtiendo para encontrar V, y luego extraer el elemento V_{00} correspondiente a la varianza de μ. La segunda derivada de $\ln \mathcal{L}$ es

$$\frac{\partial^2 \ln \mathcal{L}}{\partial\theta_j \partial\theta_k} = \sum_{i=1}^{N}\left[\left(\frac{n_i}{\nu_i}-1\right)\frac{\partial^2 \nu_i}{\partial\theta_j\partial\theta_k} - \frac{\partial \nu_i}{\partial\theta_j}\frac{\partial \nu_i}{\partial\theta_k}\frac{n_i}{\nu_i^2}\right] + \sum_{i=1}^{M}\left[\left(\frac{m_i}{u_i}-1\right)\frac{\partial^2 u_i}{\partial\theta_j\partial\theta_k} - \frac{\partial u_i}{\partial\theta_j}\frac{\partial u_i}{\partial\theta_k}\frac{m_i}{u_i^2}\right] \tag{28}$$

De (29) la segunda derivada de $\ln \mathcal{L}$ es lineal en n_i y m_i. Por lo tanto, su valor esperado se encuentra mediante la evaluación de los valores esperados de los datos, que es el mismo en los datos Asimov. Por lo tanto, se puede obtener la matriz inversa de covarianza mediante

$$V_{jk}^{-1} = -E\left[\frac{\partial^2 \ln \mathcal{L}}{\partial\theta_j\partial\theta_k}\right] = -\frac{\partial^2 \ln \mathcal{L}_A}{\partial\theta_j\partial\theta_k} = \sum_{i=1}^{N}\frac{\partial \nu_i}{\partial\theta_j}\frac{\partial \nu_i}{\partial\theta_k}\frac{1}{\nu_i} + \sum_{i=1}^{M}\frac{\partial u_i}{\partial\theta_j}\frac{\partial u_i}{\partial\theta_k}\frac{1}{u_i}. \tag{29}$$

Las derivadas de $\ln \mathcal{L}_A$ se evalúan numéricamente para encontrar la matriz inversa de la covarianza, y luego invertir y extraer la varianza de μ. De la ecuación (30) se ve que esta variación depende de los valores de los parámetros supuestos para el conjunto de datos Asimov, en particular en el parámetro asumido μ', que entra a través de la ecuación (24).

Otro método para la estimación de σ (denotado σ_A en esta sección para distinguirlo del enfoque anterior basado en las segundas derivadas de $\ln \mathcal{L}$) es encontrar el valor que es necesario para recuperar las propiedades conocidas de $-\lambda_A(\mu)$. Porque el conjunto de datos Asimov correspondiente a μ' que da $\widehat{\mu} = \mu'$, a partir de la ecuación (19) se encuentra

$$-2\ln\lambda_A(\mu) \approx \frac{(\mu - \widehat{\mu})^2}{\sigma^2} = \Lambda. \tag{30}$$

Es decir, a partir del conjunto de datos Asimov se obtiene una estimación del parámetro de no centralidad Λ que caracteriza la distribución $f(q_\mu|\mu')$. De manera equivalente, al obtener la varianza σ^2 que caracteriza la distribución de $\widehat{\mu}$ se utiliza la ecuación (31), es decir,

$$\sigma_A^2 = \frac{(\mu - \mu')^2}{q_{\mu,A}}, \tag{31}$$

donde $q_{\mu,A} = -2\ln\lambda_A(\mu)$. En el caso de exclusión de la significancia de la mediana para la hipótesis μ al suponer que no hay señal, entonces se tiene $\mu' = 0$ y por lo tanto

$$\sigma_A^2 = \frac{\mu^2}{q_{\mu,A}}, \tag{32}$$

y para la prueba estadística q_μ se establece la relación análoga. Para el caso de descubrimiento donde $\mu = 0$ se tiene,

$$\sigma_A^2 = \frac{\mu'^2}{q_{0,A}}. \tag{33}$$

Los dos métodos para obtener σ y Λ; a partir de la matriz de información de Fisher o de $q_{\mu,\Lambda}$, no son idénticos, pero se han encontrados para proporcionar resultados similares en ejemplos de interés práctico. En varios casos considerados, la distribución basada en σ_A proporciona una mejor aproximación a la verdadera distribución de muestreo que el método estándar basado en la matriz de información de Fisher, lo que lleva a la conjetura de que se incorporaran algunos de los términos de orden superior en la ecuación (31).

De manera cualitativa, bajo el supuesto de la aproximación de Wald, las pruebas estadísticas q_0, q_μ y $\widetilde{q}_\mu$ estan relacionadas con $\widehat{\mu}$, y por lo tanto sus valores de la mediana se encuentran directamente mediante el uso de la mediana de $\widehat{\mu}$, que es μ'.

4.3. Distribución de t_μ

La prueba estadística $t_\mu = -2\ln\lambda(\mu)$ de la sección 3.1 se toma como la base de la prueba estadística de un valor hipotético de μ. Esto es una prueba de $\mu = 0$ para establecer la existencia de un proceso de la señal, o los valores no nulos de μ para la obtención de un intervalo de confianza. Para el valor p de p_μ se requiere la función de densidad de probabilidad $f(t_\mu|\mu)$, y para encontrar el valor p de la mediana asumiendo un parámetro diferente necesitamos a $f(t_\mu|\mu')$.

La pdf $f(t_\mu|\mu')$ está dada por la ecuación (21), es decir,

$$f(t_\mu|\mu') = \frac{1}{2\sqrt{t_\mu}}\frac{1}{\sqrt{2\pi}}\left\{\exp\left[-\frac{1}{2}\left(\sqrt{t_\mu} + \frac{\mu-\mu'}{\sigma}\right)^2\right] + \exp\left[-\frac{1}{2}\left(\sqrt{t_\mu} - \frac{\mu-\mu'}{\sigma}\right)^2\right]\right\}. \tag{34}$$

El caso especial $\mu = \mu'$ es simplemente una distribución chi-cuadrada con un grado de libertad:

$$f\left(t_\mu|\mu'\right) = \frac{1}{\sqrt{t_\mu}}\frac{1}{\sqrt{2\pi}}e^{-t_\mu/2}. \tag{35}$$

La distribución acumulativa de t_μ asumiendo μ' es

$$F\left(t_\mu|\mu'\right) = \Phi\left(\sqrt{t_\mu} + \frac{\mu-\mu'}{\sigma}\right) + \Phi\left(\sqrt{t_\mu} - \frac{\mu-\mu'}{\sigma}\right) - 1, \tag{36}$$

donde Φ es la distribución acumulativa de la distribución Gaussiana estándar (media cero, varianza unitaria). El caso especial $\mu = \mu'$ es

$$F\left(t_\mu|\mu\right) = 2\Phi\left(\sqrt{t_\mu}\right) - 1. \tag{37}$$

El valor p de un valor hipotético μ para un valor observado t_μ sería de la manera siguiente

$$p_\mu = 1 - F\left(t_\mu|\mu\right) = 2\left[1 - \Phi\left(\sqrt{t_\mu}\right)\right], \tag{38}$$

y la significancia correspondiente es

$$Z_\mu = \Phi^{-1}\left(1 - p_\mu\right) = \Phi^{-1}\left[2\Phi\left(\sqrt{t_\mu}\right) - 1\right]. \tag{39}$$

Si el valor p se encuentra por debajo de un determinado umbral α (a menudo se toma $\alpha = 0.05$), entonces el valor de μ se rechazado o excluido en un nivel de confianza (CL) de $1-\alpha$. El conjunto de puntos no rechazados forma un intervalo de confianza con $CL = 1-\alpha$. Aquí los extremos del intervalo se obtienen al tomar $p_\mu = \alpha$ y resolviendo para μ. Al suponer la aproximación de Wald (19) y el uso de la ecuación (39) se encuentra

$$\mu_{up/lo} = \widehat{\mu} \pm \sigma\Phi^{-1}\left(1 - \frac{\alpha}{2}\right). \tag{40}$$

Una sutileza con esta fórmula es que σ depende en un cierto nivel de μ. En la práctica para encontrar los límites superior e inferior se resuelve numéricamente para valores de μ que satisfacen $p_\mu = \alpha$.

4.4. Distribución de $\widetilde{t}_\mu$

Al asumir la aproximación de Wald [2], la prueba estadística $\widetilde{t}_\mu$ en la ecuación (13) se escribe de la forma siguiente

$$\widetilde{t}_\mu = \begin{cases} \frac{\mu^2 - 2\mu\widehat{\mu}}{\sigma^2}, & \widehat{\mu} < 0, \\ \frac{(\mu-\widehat{\mu})^2}{\sigma^2}, & \widehat{\mu} \geq 0. \end{cases} \tag{41}$$

De la ecuación (42) la pdf $f\left(t_\mu|\mu'\right)$ es

$$f\left(\widetilde{t}_\mu|\mu'\right) = \frac{1}{2\sqrt{\widetilde{t}_\mu}}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\sqrt{\widetilde{t}_\mu} + \frac{\mu-\mu'}{\sigma}\right)^2\right] + \begin{cases} \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\widetilde{t}_\mu}}\exp\left[-\frac{1}{2}\left(\sqrt{\widetilde{t}_\mu} - \frac{\mu-\mu'}{\sigma}\right)^2\right] & \widetilde{t}_\mu \leq \mu^2/\sigma^2, \\ \frac{1}{\sqrt{2\pi}(2\mu/\sigma)}\exp\left[-\frac{1}{2}\frac{\left(\widetilde{t}_\mu - \frac{\mu^2-2\mu\mu'}{\sigma^2}\right)^2}{(2\mu/\sigma)^2}\right] & \widetilde{t}_\mu > \mu^2/\sigma^2. \end{cases} \tag{42}$$

En virtud del caso especial $\mu = \mu'$, la ecuación (43) se transforma en la expresión siguiente

$$f\left(\widetilde{t}_\mu|\mu'\right) = \begin{cases} \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\widetilde{t}_\mu}}e^{-\widetilde{t}_\mu/2} & \widetilde{t}_\mu \leq \mu^2/\sigma^2, \\ \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\widetilde{t}_\mu}}e^{-\widetilde{t}_\mu/2} + \frac{1}{\sqrt{2\pi}(2\mu/\sigma)}\exp\left[-\frac{1}{2}\frac{\left(\widetilde{t}_\mu + \mu^2/\sigma^2\right)^2}{(2\mu/\sigma)^2}\right] & \widetilde{t}_\mu > \mu^2/\sigma^2. \end{cases} \tag{43}$$

La correspondiente distribución acumulativa es

$$F\left(\widetilde{t}_\mu|\mu'\right) = \Phi\left(\sqrt{\widetilde{t}_\mu} + \frac{\mu-\mu'}{\sigma}\right) + \begin{cases} \Phi\left(\sqrt{\widetilde{t}_\mu} - \frac{\mu-\mu'}{\sigma}\right) - 1 & \widetilde{t}_\mu \leq \mu^2/\sigma^2, \\ \Phi\left[\frac{\widetilde{t}_\mu - \left(\mu^2 - 2\mu\mu'\right)/\sigma^2}{2\mu/\sigma}\right] - 1 & \widetilde{t}_\mu > \mu^2/\sigma^2. \end{cases} \tag{44}$$

Para $\mu = \mu'$ se tiene

$$F\left(\widetilde{t}_\mu|\mu\right) = \begin{cases} 2\Phi\left(\sqrt{\widetilde{t}_\mu}\right) - 1 & \widetilde{t}_\mu \leq \mu^2/\sigma^2, \\ \Phi\left(\sqrt{\widetilde{t}_\mu}\right) + \Phi\left(\frac{\widetilde{t}_\mu + \mu^2/\sigma^2}{2\mu/\sigma}\right) - 1 & \widetilde{t}_\mu > \mu^2/\sigma^2. \end{cases} \tag{45}$$

El valor p de la hipótesis μ es

$$p_\mu = 1 - F\left(\widetilde{t}_\mu|\mu\right). \qquad (46)$$

La correspondiente significancia es $Z_\mu = \Phi^{-1}\left(1 - p_\mu\right)$.

Un intervalo de confianza para μ con un nivel de confianza $CL = 1 - \alpha$ se construye a partir de los valores de conjunto μ para los que el valor p no es menor que α. Los puntos extremos de este intervalo se encuentran al establecer p_μ de la ecuación (47) igual a α y resolver para μ. En general, $p_\mu = \alpha$ se resuelve numéricamente. En el límite de la muestra grande, es decir, al suponer la validez de las aproximaciones asintóticas, estos intervalos corresponden a los límites de Feldman y Cousins [12] para el caso $\mu \geq 0$.

4.5. Distribución de q_0 (descubrimiento)

En virtud de la aproximación (19) se tiene que $-2\ln\lambda(0) = \widehat{\mu}^2/\sigma^2$. De la ecuación (14) para q_0 se tiene

$$q_0 = \begin{cases} \widehat{\mu}^2/\sigma^2 & \widehat{\mu} \geq 0 \\ 0 & \widehat{\mu} < 0., \end{cases} \qquad (47)$$

donde $\widehat{\mu}$ sigue una distribución Gaussiana con media μ' y desviación estándar σ. La pdf de q_0 tiene la forma

$$f\left(q_0|\mu'\right) = \left[1 - \Phi\left(\frac{\mu'}{\sigma}\right)\right]\delta\left(q_0\right) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}\exp\left[-\frac{1}{2}\left(\sqrt{q_0} - \frac{\mu'}{\sigma}\right)^2\right]. \qquad (48)$$

Para el caso especial de $\mu' = 0$, la ecuación (49) se reduce a

$$f\left(q_0|\mu' = 0\right) = \delta\left(q_0\right) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}e^{-q_0/2}. \qquad (49)$$

Es decir, se encuentra una mezcla de una función delta en cero y una distribución chi-cuadrada con un grado de libertad. En el resto del capítulo nos referiremos a esta mezcla como una distribución chi-cuadrada media o $\frac{1}{2}\chi^2$.

De la ecuación (49) la distribución acumulativa correspondiente sería

$$F\left(q_0|\mu'\right) = \Phi\left(\sqrt{q_0} - \frac{\mu'}{\sigma}\right). \qquad (50)$$

El caso especial $\mu' = 0$ es

$$F\left(q_0|0\right) = \Phi\left(\sqrt{q_0}\right). \qquad (51)$$

El valor p de la hipótesis $\mu = 0$ es de la forma

$$p_0 = 1 - F\left(q_0|0\right), \qquad (52)$$

y por lo tanto al usar la ecuación (3) para la significancia es

$$Z_0 = \Phi^{-1}\left(1 - p_0\right) = \sqrt{q_0}. \qquad (53)$$

4.6. Distribución de q_μ

Mediante la aproximación de Wald se escribe la prueba estadística utilizada para los límites superiores, la ecuación (16) da

$$q_\mu = \begin{cases} \frac{(\mu - \widehat{\mu})^2}{\sigma^2} & \widehat{\mu} < \mu \\ 0 & \widehat{\mu} > \mu, \end{cases} \qquad (54)$$

donde $\widehat{\mu}$ como antes sigue una distribución Gaussiana centrada sobre μ' con una desviación estándar σ.

La pdf $f\left(q_\mu|\mu'\right)$ es

$$f\left(q_\mu|\mu'\right) = \Phi\left(\frac{\mu - \mu'}{\sigma}\right)\delta\left(q_\mu\right) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_\mu}}\exp\left[-\frac{1}{2}\left(\sqrt{q_\mu} - \frac{\mu - \mu'}{\sigma}\right)^2\right], \qquad (55)$$

así que el caso especial $\mu = \mu'$ es la distribución chi-cuadrada media

$$f\left(q_\mu|\mu'\right) = \frac{1}{2}\delta\left(q_\mu\right) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_\mu}}e^{-\frac{1}{2}q_\mu}. \qquad (56)$$

La distribución acumulativa se expresa mediante

$$F\left(q_\mu|\mu'\right) = \Phi\left(\sqrt{q_\mu} - \frac{\mu - \mu'}{\sigma}\right), \qquad (57)$$

y el caso $\mu' = \mu$ sería

$$F\left(q_\mu|\mu\right) = \Phi\left(\sqrt{q_\mu}\right). \qquad (58)$$

El valor p del hipotético μ es

$$p_\mu = 1 - F\left(q_\mu|\mu\right) = 1 - \Phi\left(\sqrt{q_\mu}\right), \qquad (59)$$

y, por lo tanto para la significancia

$$Z_\mu = \Phi^{-1}\left(1 - p_\mu\right) = \sqrt{q_\mu}. \qquad (60)$$

Si se encuentra que el valor p por debajo de un umbral α especificado (a menudo se toma $\alpha = 0.05$), entonces el valor de μ se rechaza para un nivel de confianza (CL) de $1 - \alpha$. El límite superior de μ es la más grande μ con $p_\mu \leq \alpha$. Aquí esto se puede obtener simplemente fijando $p_\mu = \alpha$ y resolviendo para μ. Al las ecuaciones (55) y (50) se encuentra

$$\mu_{up} = \widehat{\mu} + \sigma\Phi^{-1}\left(1 - \alpha\right). \qquad (61)$$

Por ejemplo, $\alpha = 0.05$ da $\Phi^{-1}\left(1 - \alpha\right) = 1.64$. También, σ depende de μ. Por lo tanto, el límite superior se encuentra numéricamente como el valor de μ para el cual $p_\mu = \alpha$.

4.7. Distribución de $\widetilde{q}_\mu$

La prueba estadística alternativa $\widetilde{q}_\mu$ definida en la ecuación (18) y al asumir la aproximación de Wald se escribe mediante

$$\widetilde{q}_\mu = \begin{cases} \frac{\mu^2 - 2\mu\widehat{\mu}}{\sigma^2} & \widehat{\mu} < 0, \\ \frac{(\mu - \widehat{\mu})^2}{\sigma^2} & 0 \le \widehat{\mu} \le \mu, \\ 0 & \widehat{\mu} > \mu. \end{cases} \tag{62}$$

La pdf $f\left(\widetilde{q}_\mu|\mu'\right)$ encontrada es

$$f\left(\widetilde{q}_\mu|\mu'\right) = \Phi\left(\frac{\mu' - \mu}{\sigma}\right)\delta\left(\widetilde{q}_\mu\right) + \begin{cases} \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\widetilde{q}_\mu}}\exp\left[-\frac{1}{2}\left(\sqrt{\widetilde{q}_\mu} - \frac{\mu-\mu'}{\sigma}\right)^2\right] & 0 < \widetilde{q}_\mu \le \mu^2/\sigma^2, \\ \frac{1}{\sqrt{2\pi}(2\mu/\sigma)}\exp\left[-\frac{1}{2}\frac{\left(\widetilde{q}_\mu - (\mu^2 - 2\mu\mu')/\sigma^2\right)^2}{(2\mu/\sigma)^2}\right] & \widetilde{q}_\mu > \mu^2/\sigma^2. \end{cases} \tag{63}$$

El caso especial $\mu = \mu'$ es por lo tanto

$$f\left(\widetilde{q}_\mu|\mu\right) = \frac{1}{2}\delta\left(\widetilde{q}_\mu\right) + \begin{cases} \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\widetilde{q}_\mu}}e^{-\widetilde{q}_\mu/2} & 0 < \widetilde{q}_\mu \le \mu^2/\sigma^2, \\ \frac{1}{\sqrt{2\pi}(2\mu/\sigma)}\exp\left[-\frac{1}{2}\frac{\left(\widetilde{q}_\mu + \mu^2/\sigma^2\right)^2}{(2\mu/\sigma)^2}\right] & \widetilde{q}_\mu > \mu^2/\sigma^2. \end{cases} \tag{64}$$

La correspondiente distribución acumulativa es

$$F\left(\widetilde{q}_\mu|\mu'\right) = \begin{cases} \Phi\left(\sqrt{\widetilde{q}_\mu} - \frac{\mu-\mu'}{\sigma}\right) & 0 < \widetilde{q}_\mu \le \mu^2/\sigma^2, \\ \Phi\left(\frac{\widetilde{q}_\mu - (\mu^2 - 2\mu\mu')/\sigma^2}{2\mu/\sigma}\right) & \widetilde{q}_\mu > \mu^2/\sigma^2. \end{cases} \tag{65}$$

El caso especial $\mu = \mu'$ sería

$$F\left(\widetilde{q}_\mu|\mu\right) = \begin{cases} \Phi\left(\sqrt{\widetilde{q}_\mu}\right) & 0 < \widetilde{q}_\mu \le \mu^2/\sigma^2, \\ \Phi\left(\frac{\widetilde{q}_\mu + \mu^2/\sigma^2}{2\mu/\sigma}\right) & \widetilde{q}_\mu > \mu^2/\sigma^2. \end{cases} \tag{66}$$

El valor p de μ se expresa mediante

$$p_\mu = 1 - F\left(\widetilde{q}_\mu|\mu\right), \tag{67}$$

y por lo tanto la significancia es

$$Z_\mu = \begin{cases} \sqrt{\widetilde{q}_\mu} & 0 < \widetilde{q}_\mu \le \mu^2/\sigma^2, \\ \frac{\widetilde{q}_\mu + \mu^2/\sigma^2}{2\mu/\sigma} & \widetilde{q}_\mu > \mu^2/\sigma^2. \end{cases} \tag{68}$$

El límite superior de μ con un nivel de confianza $1-\alpha$ se encuentra al establecer $p_\mu = \alpha$ y resolviendo para μ, lo que reduce al mismo resultado de q_μ, es decir,

$$\mu_{up} = \widehat{\mu} + \sigma\Phi^{-1}\left(1 - \alpha\right). \tag{69}$$

Si la aproximación de Wald se mantiene, entonces las dos pruebas estadísticas q_μ y $\widetilde{q}_\mu$ conducen a límites superiores idénticos.

4.8. Distribución de $-2\ln\left(\mathcal{L}_{s+b}/\mathcal{L}_b\right)$

Muchos análisis del colisionador Tevatron (por ejemplo, [14]) que implican búsquedas nuevas para procesos de señal se han basado en la prueba estadística

$$q = -2\ln\left(\frac{\mathcal{L}_{s+b}}{\mathcal{L}_b}\right), \tag{70}$$

donde $\mathcal{L}_{s+b}$ es la verosimilitud del modelo de señal nominal y $\mathcal{L}_b$ es la verosimilitud de la hipótesis de fondo. Es decir, $s+b$ corresponde $\mu = 1$ y $\mathcal{L}_b$ se refiere a $\mu = 0$. La prueba estadística q por lo tanto se escribe de la forma

$$q = -2\ln\frac{\mathcal{L}\left(\mu = 1, \widehat{\widehat{\theta}}\left(1\right)\right)}{\mathcal{L}\left(\mu = 0, \widehat{\widehat{\theta}}\left(0\right)\right)} = -2\ln\lambda\left(1\right) + 2\ln\lambda\left(0\right). \tag{71}$$

Al asumir la aproximación de Wald (29), q está dada por

$$q = \frac{(\widehat{\mu} - 1)^2 - \widehat{\mu}^2}{\sigma^2} = \frac{1 - 2\widehat{\mu}}{\sigma^2}, \tag{72}$$

donde σ^2 es la varianza de μ. Como μ sigue una distribución Gaussiana, la distribución de q también es Gaussiana, con un valor medio de

$$E\left[q\right] = \frac{1 - 2\mu}{\sigma^2}, \tag{73}$$

y una varianza de

$$V\left[q\right] = \frac{4}{\sigma^2}. \tag{74}$$

Es decir, la desviación estándar de q es $\sigma_q = \sqrt{V\left[q\right]} = 2/\sigma$, donde la desviación estándar de $\widehat{\mu}$ es σ, y se estima al utilizar la segunda derivada de la función de verosimilitud logarítmica descrita en la sección 4.1 o con los métodos descritos en la sección 4.2. σ en general depende de μ; aquí nos referiremos a estos como σ_b y σ_{s+b} para $\mu = 0$ y $\mu = 1$, respectivamente.

De la ecuación (74) en la hipótesis $s+b$ ($\mu = 1$) los valores de q tienden a ser más bajos, y para la hipótesis

b $(\mu = 0)$ son más altos. Por lo tanto se encuentran los valores p para las dos hipótesis a partir de las expresiones

$$p_{s+b} = \int_{q_{obs}}^{\infty} f\left(q|s+b\right) dq = 1 - \Phi\left(\frac{q_{obs} + 1/\sigma_{s+b}^2}{2/\sigma_{s+b}}\right), \tag{75}$$

$$p_b = \int_{-\infty}^{q_{obs}} f\left(q|s\right) dq = \Phi\left(\frac{q_{obs} - 1/\sigma_b^2}{2/\sigma_b}\right), \tag{76}$$

donde se utilizaron las ecuaciones (74) y (75) para la media y la varianza de q con las hipótesis b y $s + b$.

Los valores p de las ecuaciones (76) y (77) incorporan los efectos de las incertidumbres sistemáticas con los parámetros de perturbación $\overrightarrow{\theta}$. Los efectos de las incertidumbres sistemáticas se tienen en cuenta mediante el uso de las mediciones de control como la base de una densidad Bayesiana a priori $\pi\left(\overrightarrow{\theta}\right)$, y la distribución de q se calcula con el promedio Bayesiano [15]

$$f\left(q\right) = \int f\left(q|\overrightarrow{\theta}\right) \pi\left(\overrightarrow{\theta}\right) d\overrightarrow{\theta}. \tag{77}$$

La pdf $\pi\left(\overrightarrow{\theta}\right)$ utilizada en la ecuación (78) se obtiene a partir de algunas de las mediciones que caracterizan una función de verosimilitud $\mathcal{L}_{\overrightarrow{\theta}}\left(\overrightarrow{\theta}\right)$, y luego usarla para encontrar $\pi\left(\overrightarrow{\theta}\right)$ usando el teorema de Bayes [16–18],

$$\pi\left(\overrightarrow{\theta}\right) \propto \mathcal{L}_{\overrightarrow{\theta}}\left(\overrightarrow{\theta}\right) \pi_0\left(\overrightarrow{\theta}\right). \tag{78}$$

Aquí $\pi_0\left(\overrightarrow{\theta}\right)$ es la pdf inicial a priori. Esta pdf refleja conocimientos antes de llevar a cabo las mediciones de control. En muchos casos, se toma como una constante, en cuyo caso $\pi\left(\overrightarrow{\theta}\right)$ es simplemente proporcional a $\mathcal{L}_{\overrightarrow{\theta}}\left(\overrightarrow{\theta}\right)$.

4.9. Sensibilidad experimental

Al caracterizar la sensibilidad de un experimento, se está interesado no en el significado obtenido a partir de un único conjunto de datos, sino más bien en el significancia esperada (más precisamente, la mediana) con la que se rechazan valores diferentes de μ. En concreto, para el caso del descubrimiento nos gustaría conocer la mediana, con del modelo de señal nominal $(\mu = 1)$, con la que se podría rechazar la hipótesis de fondo $(\mu = 0)$.

La sensibilidad de un experimento se ilustra en la figura 2, la cual muestra la pdf para q_μ al asumir tanto un parámetro μ como un valor μ' diferente. La distribución $f\left(q_\mu|\mu'\right)$ se desplaza a mayor valor de q_μ, lo que corresponde en promedio a bajar los valores p. La sensibilidad de un experimento se caracteriza al dar el valor p correspondiente a la mediana q_μ suponiendo el valor alternativo μ'. A medida que el valor p es una función de q_μ, esto es igual al valor p de la mediana asumiendo μ'.

En el resto de esta sección se describen los ingredientes necesarios para determinar la sensibilidad experimental (descubrimiento de la mediana o la significancia de rechazo).

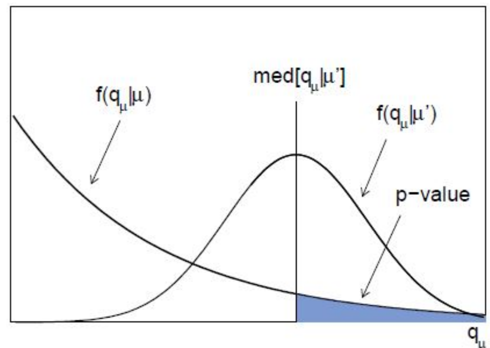

Figura 2: Ilustración del valor p correspondiente a la mediana de q_μ asumiendo un parámetro μ'.

4.10. La significancia de la mediana de valores Asimov en la prueba estadística

Al utilizar el conjunto de datos Asimov se obtienen los valores de la mediana de q_0, q_μ y $\widetilde{q}_\mu$, y ello conduce a expresiones simples para la correspondiente significancia de la mediana. A partir de las ecuaciones (54), (61) y (69) el significado de Z es una función de q, y por lo tanto la mediana Z está dada por la función de la mediana de q. Para el descubrimiento al usar q_0 se quiere la "median discovery significance" al asumir un parámetro μ' y para los límites superiores se está interesado en la zona de rechazo de la significancia de la mediana al suponer $\mu' = 0$, med $\left[Z_\mu|0\right]$. Para estos valores se obtiene

$$\text{med}\left[Z_0|\mu'\right] = \sqrt{q_{0,A}}, \tag{79}$$

$$\text{med}\left[Z_\mu|0\right] = \sqrt{q_{\mu,A}}. \tag{80}$$

Cuando se usa $\widetilde{q}_\mu$ para establecer límites superiores, la expresión general de la significancia de exclusión Z_μ es algo más complicada dependiendo de μ', pero en cualquier caso se encuentra mediante la sustitución de los valores apropiados de $\widetilde{q}_{\mu,A}$ y σ_A en la ecuación (69). Para el caso donde se desea la significancia de la mediana para μ al asumir datos distribuidos de acuerdo a la hipótesis de fondo $(\mu = 0)$, la ecuación (69) se reduce a una relación de la misma forma como la ecuación (61), y por lo tanto se encuentra

$$\text{med}\left[Z_\mu|0\right] = \sqrt{\widetilde{q}_{\mu,A}}. \tag{81}$$

4.11. Combinación de múltiples canales

En muchos análisis, se pueden combinar varios canales de búsqueda. Para cada canal i hay una función de verosimilitud $\mathcal{L}_i\left(\mu, \overrightarrow{\theta}_i\right)$, donde $\overrightarrow{\theta}_i$ representa el conjunto de parámetros de perturbación para el canal i-ésimo, algunos de los cuales puede ser comúnes entre los canales. Aquí, el parámetro μ es el mismo para todos los canales. Si los canales son estadísticamente independientes, la función de verosimilitud completa está dada por el producto sobre la totalidad de los canales, es decir:

$$\mathcal{L}\left(\mu, \overrightarrow{\theta}\right) = \prod_i \mathcal{L}_i\left(\mu, \overrightarrow{\theta}_i\right), \tag{82}$$

donde $\overrightarrow{\theta}_i$ representa a todos los parámetros de perturbación. Por consiguiente, el cociente de verosimilitud $\lambda\left(\mu\right)$ es

$$\lambda\left(\mu\right) = \frac{\prod_i \mathcal{L}_i\left(\mu, \widehat{\widehat{\overrightarrow{\theta}}}_i\right)}{\prod_i \mathcal{L}_i\left(\widehat{\mu}, \widehat{\overrightarrow{\theta}}_i\right)}. \tag{83}$$

Debido a que los datos Asimov no contienen fluctuaciones estadísticas, se tiene $\widehat{\mu} = \mu'$ para todos los canales. Además cualquiera de los componentes comunes de $\overrightarrow{\theta}_i$ son los mismos para todos los canales. Por lo tanto, cuando se utilizan los datos Asimov correspondientes al parámetro μ' se encuentra que

$$\lambda_A\left(\mu\right) = \frac{\prod_i \mathcal{L}_i\left(\mu, \widehat{\widehat{\overrightarrow{\theta}}}\right)}{\prod_i \mathcal{L}_i\left(\mu', \overrightarrow{\theta}\right)} = \prod_i \lambda_{A,i}\left(\mu\right), \tag{84}$$

donde $\lambda_{A,i}\left(\mu\right)$ es la razón de verosimilitud Asimov sólo para el i-ésimo canal.

Debido a esto, se determinan los valores de la relación de verosimilitud al entrar en (85) por separado para cada canal, lo que simplifica la tarea de estimar la significancia de la mediana que resultaría de la combinación completa. Sin embargo, para encontrar la significancia del descubrimiento o de rechazo determinada a partir de datos reales, hay que construir la función de verosimilitud completa que contiene un parámetro simple μ, y esto debe ser utilizado en un ajuste global para encontrar el cociente de la función de verosimilitud.

4.12. Variación estadística esperada (bandas de error)

Mediante el uso del conjunto de datos Asimov fijo podemos encontrar la mediana, al suponer algún parámetro μ de la significancia del rechazo de un valor hipotético μ. Incluso si el valor hipotético μ' es correcto, los datos reales contendrán fluctuaciones estadísticas, y por lo tanto la significancia observada no es igual a la mediana.

Por ejemplo, si la señal está, de hecho, ausente, pero el número de eventos de fondo fluctúa hacia arriba, entonces el límite superior observado en el parámetro μ será más débil que la mediana al suponer solamente el fondo. La significancia se espera que varie, una vez que se tomen en cuenta las fluctuaciones esperadas en los datos. Debido a que se tienen fórmulas para todas las distribuciones de muestreo correspondientes, también se predice cómo se espera que la significancia debe variar ante la suposición de una intensidad de señal dada.

Es conveniente el cálculo de bandas de error para la significancia de la mediana correspondiente a la variación $\pm N\sigma$ de $\widehat{\mu}$. Como μ se distribuye de forma Gaussiana, estas bandas de error sobre la significancia son simplemente los cuantiles que se asignan a la variación de $\widehat{\mu}$ de $\pm N\sigma$ acerca de μ'.

Para el caso de descubrimiento, esto es, una prueba de $\mu = 0$, de las ecuaciones (48) y (54) se tiene que la significancia Z_0 es

$$Z_0 = \begin{cases} \frac{\widehat{\mu}}{\sigma}, \ \widehat{\mu} \geq 0, \\ 0, \ \widehat{\mu} < 0. \end{cases} \tag{85}$$

Por otra parte la significancia de la mediana se calcula a partir de la ecuación (80), así los valores de significancia correspondientes a $\mu' \pm N\sigma$ son

$$Z_0\left(\mu' + N\sigma\right) = \mathrm{med}\left[Z|\mu'\right] + N, \tag{86}$$

$$Z_0\left(\mu' - N\sigma\right) = \max\left[\mathrm{med}\left[Z|\mu'\right] - N, 0\right]. \tag{87}$$

Para el caso de la exclusión (rechazo), cuando se utiliza tanto la estadística q_μ, así como $\widetilde{q}_\mu$ se encontró la misma expresión para el límite superior con un nivel de confianza de $1-\alpha$, es decir, la ecuación (62). Por lo tanto, el límite superior de la mediana al suponer un parámetro μ' se encuentra mediante la sustitución de esto para $\widehat{\mu}$, y bandas de error de $\pm N\sigma$ se encuentran de manera similar mediante la sustitución de los correspondientes valores de $\mu' \pm N\sigma$. Es decir, el límite superior de la mediana es

$$\mathrm{med}\left[\mu_{up}|\mu'\right] = \mu' + \sigma\Phi^{-1}\left(1-\alpha\right), \tag{88}$$

y las bandas de error $\pm N\sigma$ están dadas por

$$\mathrm{banda}_{N\sigma} = \mu\text{'} + \sigma\left[\Phi^{-1}\left(1-\alpha\right) \pm N\right]. \tag{89}$$

La desviación estándar σ de $\widehat{\mu}$ se obtiene de los valores Asimov de la prueba estadística q_μ (o $\widetilde{q}_\mu$) usando la ecuación (32).

5. Conclusiones

Las pruebas estadísticas en física de partículas[3] se describen para su uso en la planificación y la realización de una búsqueda de nuevos fenómenos. El formalismo permite el tratamiento de las incertidumbres sistemáticas a través del uso de la razón de funciones de verosimilitud. Aquí una incertidumbre sistemática es incluida en la medida en que el modelo incluye un número suficiente de parámetros de perturbación de manera que por lo menos en algún punto de su espacio de parámetros pueda ser considerada como verdad.

En este artículo se deducen fórmulas aproximadas para las distribuciones de las pruebas estadísticas utilizadas para caracterizar el grado de concordancia entre los datos y la hipótesis que se prueba, así como las expresiones relacionadas a los valores p y las significancias. Los estadísticos se basan en la razón de funciones de verosimilitud y se pueden utilizar para una prueba bilateral de un parámetro μ (t_μ), una prueba unilateral para el descubrimiento (q_0), y una prueba unilateral para encontrar un límite superior (q_μ y $\widetilde{q}_\mu$). El estadístico $\widetilde{t}_\mu$ se usa para obtener un intervalo de confianza "unificado", en el sentido de que es unilateral o bilateral, dependiendo del resultado de los datos.

También se proporcionan las fórmulas que permiten caracterizar la sensibilidad de un experimento planificado a través de la significancia de la mediana de una hipótesis dada bajo el supuesto de una diferente, por ejemplo, la significancia de la mediana con la que se podría rechazar la hipótesis de fondo ante el supuesto de una cierta señal de modelo. Estas explotan el uso de un conjunto artificial de datos, el conjunto de datos "Asimov", que se define a fin de que los estimadores para todos los parámetros sean iguales a sus valores verdaderos. También se dan métodos para encontrar la variación estadística esperada en la sensibilidad (bandas de error).

Apéndice: Prueba estadística $-2\ln Q\left(m_H\right)$

La prueba estadística adoptada en la combinación de datos LEP [36] es $-2\ln Q$, donde Q es la división de la función de verosimilitud para la hipótesis de “señal + fondo” y la función de verosimilitud para la hipótesis de “fondo”, es decir

$$Q=\frac{\mathcal{L}_{s+b}}{\mathcal{L}_b}. \tag{90}$$

Las funciones de verosimilitud agrupadas se definen por

$$\mathcal{L}\left(\mu\right)=\prod_{k=1}^{N}\frac{\left[\mu s_k\left(m_H\right)+b_k\right]^{n_k}\exp\left[-\left(\mu s_k\left(m_H\right)+b_k\right)\right]}{n_k!}\prod_{j=1}^{n_k}\frac{\mu s_k\left(m_H\right)S_k\left(\overrightarrow{x}_{ij};m_H\right)+b_kB_k\left(\overrightarrow{x}_{jk}\right)}{\mu s_k\left(m_H\right)+b_k}. \tag{91}$$

Donde $\mu=1$ en el caso de $\mathcal{L}_{s+b}$ y $\mu=0$ en el caso de $\mathcal{L}_b$. El índice k corre sobre todas las contribuciones independientes al resultado combinado: a partir de diferentes selecciones topológicas, los datos se toman a diferentes energías del centro de masa y los datos recogidos en diferentes experimentos. El símbolo N representa el número de tales contribuciones ("canales"); n_k es el número de candidatos observados en el canal k y $\overrightarrow{x}_{jk}$ designa la posición x del candidato j del canal k en el plano definido por las variables discriminantes m_H^{rec} y $\mathcal{G}$ (véase [37]). Las cantidades $s_k\left(m_H\right)$ y b_k son los tipos de señal y de fondo integrados en el canal k. Las funciones $S_k\left(\overrightarrow{x}_{jk};m_H\right)$ y $B_k\left(\overrightarrow{x}_{jk}\right)$ son las funciones de densidad de probabilidad (PDF) de las variables discriminantes para la señal y el fondo. Estos PDFs se evalúan en intervalos de m_H^{rec} y $\mathcal{G}$ de un conjunto de valores para m_H con alguna interpolación y procedimientos de suavizado aplicado [38, 39]. La prueba estadística se puede escribir como

$$-2\ln Q\left(m_H\right)=2\sum_{k=1}^{N}\left[s_k\left(m_H\right)-\sum_{j=1}^{n_k}\ln\left(1+\frac{s\left(m_H\right)S_k\left(\overrightarrow{x}_{jk},m_H\right)}{b_kB_k\left(\overrightarrow{x}_{jk}\right)}\right)\right], \tag{92}$$

Así se convirte en una suma de contribuciones (pesos) de los eventos observados individuales. La notación de la ecuación precedente asume que las cantidades de fondo relacionadas, es decir, b_k y $B_k\left(\overrightarrow{x}_{jk}\right)$ no dependen de m_H. En el citerio de selección en alguno de los canales son explicitamente dependientes de m_H (las investigaciones de L3 y OPAL en el cuarto canal tienen esta propiedad), b_k y $B_k\left(\overrightarrow{x}_{jk}\right)$ se reemplazan por , $b_k\left(m_H\right)$ y $B_k\left(\overrightarrow{x}_{jk};m_H\right)$.

La presencia de una señal se infiere a partir del comportamiento de la confianza $1-CL_B$ para la hipótesis de fondo (también llamado valor p, véase Ref. [40]), la cual es obtenida, por una masa de prueba dada, median-

[3] Se fundamenta en el formalismo Lagrangiano de las teorías de campos [19–35].

te la integración de la correspondiente PDF desde $-\infty$ al valor observado de la prueba estadística. Las PDFs se obtienen a partir de simulaciones detalladas de experimentos, dada la hipótesis de fondo. Si la hipótesis de fondo es correcta, $1 - CL_b$ se distribuye uniformemente entre cero y uno; la mediana de la distribución sería 0.5. En presencia de una señal significativa $1-CL_b$ sería muy pequeño para la masa de prueba correspondiente.

El límite de rechazo (exclusión) frecuencial por lo general se calcula a partir de CL_{s+b} para la hipótesis de la señal más el fondo la cual, por una masa de prueba dada, se obtiene integrando la correspondiente PDF con el valor observado de la prueba estadística para $+\infty$. La hipótesis de la señal más el fondo se rechaza al nivel de confianza del 95 % si la observación se hace tal que CL_{s+b} es menor que 0.05. Sin embargo, este procedimiento puede dar lugar a la posibilidad no deseada de que una gran fluctuación hacia abajo del fondo permitiría a la hipótesis ser rechazada para la cual el experimento no tenga ninguna sensibilidad debido a la pequeña rapidez de la señal esperada. Este problema se evita mediante la introducción de la relación de $CL_S = CL_{s+b}/CL_b$. Ya que CL_b es un número positivo menor que uno, CL_s siempre será mayor que CL_{s+b}, y el límite obtenido será conservativo. Se adopta esta cantidad para establecer los límites de rechazo y se considera una hipótesis que debe rechazarse en el nivel de confianza del 95 % si el valor correspondiente de CL_s es inferior a 0.05.

Referencias

[1] S. S. Wilks, The large-sample distribution of the likelihood ratio for testing composite hypotheses, Annals of Mathematical Statistics, 9, 60-2 (1938).

[2] A. Wald, Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large, Transactions of the American Mathematical Society, 54, 426-482 (1943).

[3] P. A. Zyla et al. (Particle Data Group), Review of Particle Physics, Progress of Theoretical and Experimental Physics, 8, 083C01 (2020).

[4] R. L. Workman, et al. (Particle Data Group), Review of Particle Physics, Progress of Theoretical and Experimental Physics, 8, 083C01 (2022).

[5] V. Bartsch and G. Quast, Expected signal observability at future experiments, CMS Note 2005/004 (2003) (available on CMS information server).

[6] ATLAS Collaboration, Expected Performance of the ATLAS Experiment, Detector, Trigger and Physics, CERN-OPEN-2008-020, Geneva, (2008), ⟨ArXiv:hep-ex/0901.0512⟩.

[7] M. Valenzuela, Pruebas Estadísticas en la Física del bosón de Higgs, Tesis de Maestría en Matemáticas Aplicadas, Universidad Autónoma de Zacatecas, Mexico, february 2018.

[8] ALEPH, DELPHI, L3 and OPAL Collaborations, Search for the Standard Model Higgs Boson at LEP, CERN-EP/2003-011, Physical Letters B, 565, 61-75 (2003).

[9] Eric Burns, Wade Fisher, Testing the approximations described in “Asymptotic formulae for likelihood-based tests of new physics”, 2011, ⟨ArXiv:hep-ex/1110.5002⟩.

[10] G. Cowan, K. Cranmer, E. Gross y O. Vitells, Asymptotic formulae for likelihood-based tests of new physics, Journal of European Physics C, 71, 1554 (2011).

[11] A. Stuart, J. K. Ord, y S. Arnold, Kendalls Advanced Theory of Statistics, Vol. 2A: Classical Statistical Inference and the Linear Model, sixth Edition (Oxford University Press, 1999), and earlier editions by Kendall and Stuart.

[12] Robert D. Cousins and Gary J. Feldman, A Unified Approach to the Classical Statistical Analysis of Small Signals, Physical Review D, 57, 3873 (1998).

[13] M. Abramowitz y I. A. Stegun, Handbook of Mathematical Functions, Sección 26.4.25, Dover, USA (1972).

[14] T. Aaltonen et al., Combination of Tevatron Searches for the Standard Model Higgs Boson in the W^+W^- Decay Mode, Physical Review Letters, 104, 061802 (2010).

[15] G. Cowan, Statistical Data Analysis. Oxford University Press Inc., New York USA (1998).

[16] G. Cowan, Statistics for Searches at the LHC, 69th Scottish Universities Summer School in Physics, St. Andrews, August-September 2012, ⟨ArXiv:hep-ex/1307.2487⟩.

[17] G. Cowan, Topics in statistical data analysis for high-energy physics, T. Binoth, C. Buttar, P. J. Clarck, E. W. N. Glover (eds.), Taylor y Francis, 2012, ⟨ArXiv:1012.3589v1 [physics.data-an]⟩.

[18] K. Cranmer, Practical Statistics for the LHC, Proceedings of the 2013 CERN–Latin-American School of High-Energy Physics, Arequipa, Peru, 6-19 March 2013, edited by M. Mulders and G. Perez, CERN-2015-001 (CERN, Geneva, 2015), ⟨ArXiv:1503.07622v1[physics.data-an]⟩.

[19] Thomson, M., Modern Particle Physics, Cambridge University Press (2013).

[20] Greenwood, D. A. y Cottingham, W. N, An introduction to the standard model of particle physics. Cambridge University Press (2007).

[21] Maggiore, M., A Modern Introduction to Quantum Field Theory, Oxford University Press (2005).

[22] Weinberg, S., Quantum theory of fields. Foundations, volumen 1, Cambridge University Press (1995).

[23] Pokorski, S., Gauge Field Theories, Cambridge University Press (2000).

[24] M. Valenzuela, Clasificación algebraica de los modelos de Bianchi, Tesis de Licenciatura en Física, Universidad Autónoma de Zacatecas, México, (2012).

[25] M. Valenzuela, Hamiltonian formalism of the Bianchi's models, Revista de Investigación de Física, 25(1), 26-48 (2022). Doi: https://doi.org/10.15381/rif.v25i1.21413

[26] M. Valenzuela, A relativistic theory of the field, Revista de Investigación de Física, 24(2), 72-79 (2021). Doi: https://doi.org/10.15381/rif.v24i2.14245

[27] M. Valenzuela, A relativistic theory of the field. II: Hamilton's principle and Bianchi's identities, Revista de Investigación de Física, 24(3), 12-24 (2021). Doi: https://doi.org/10.15381/rif.v24i3.14375

[28] M. Valenzuela, General relativistic theory of gravitation and electrodynamics, Revista de Investigación de Física, 25(1), 1-9 (2022). Doi: https://doi.org/10.15381/rif.v25i1.14972

[29] M. Valenzuela, Corrigendum: A relativistic theory of the field, Revista de Investigación de Física, 25(2), 35-36 (2022). Doi: https://doi.org/10.15381/rif.v25i2.22486

[30] M. Valenzuela, Corrigendum: A relativistic theory of the field II, Revista de Investigación de Física, 25(2), 37-38 (2022). Doi: https://doi.org/10.15381/rif.v25i2.22487

[31] M. Valenzuela, On the relativistic theory of the asymmetric field, Revista de Investigación de Física, 25(2), 21-34 (2022). Doi: https://doi.org/10.15381/rif.v25i2.21449

[32] M. Valenzuela, "Corrigendum: On the relativistic theory of the asymmetric field", Revista de Investigación de Física, 25(3), 45-47 (2022). Doi: https://doi.org/10.15381/rif.v25i3.23719

[33] M. Valenzuela, Gravitation, cosmology and dark matter, Revista de Investigación de Física, 26(1), 35-68 (2023). Doi: https://doi.org/10.15381/rif.v26i1.22806

[34] M. Valenzuela, Global symmetry breaking and topological defects, Revista de Investigación de Física, 26(2), 25-38 (2023). Doi: https://doi.org/10.15381/rif.v26i2.23688

[35] M. Valenzuela, Relativistic theory of gravitation for Modified Newtonian Dynamics, Revista de Investigación de Física, 26(3), 1-5 (2023). Doi: https://doi.org/10.15381/rif.v26i3.23641

[36] T. Junk, Confidence level computation for combining searches with small statistics, Nuclear Instruments, Nuclear Instruments and Methods in Physics Research **A**, **434** (1999) 435-443. A. L. Read, "Presentation of search results: the CLs technique", Journal of Physics **G**, **28** 2693 (2002).

[37] ALEPH, DELPHI, L3 y OPAL Collaborations, Search for the Standard Model Higgs Boson at LEP, CERN-EP/2003-011, Physics Letters B, 565, 61-75 (2003). ⟨ArXiv:hep-ex/0306033⟩.

[38] A.L. Read, Linear interpolation of histograms, Nuclear Instruments Methods in Physics Research A, 225, 357-360 (1999).

[39] K.S. Cranmer, Kernel estimation in high-energy physics, Computer Physics Communications, 136, 198 (2001).

[40] K. Hagiwara et al. (Particle Data Group), Review of Particle Properties, Physical Review D, 66, 010001-1, Review No. 31 on Statistics, p. 229. (2002).