

**Centro de Investigación y de Estudios Avanzados
del
Instituto Politécnico Nacional**

DEPARTAMENTO DE FÍSICA

**Aplicación de la Técnica de Feldman-Cousins
al Decaimiento $B^0 \rightarrow K_s^0 \mu^+ \mu^-$**

Tesis que presenta

Antonio Cota Rodríguez

para obtener el Grado de

Maestro en Ciencias en la Especialidad de Física

Director de tesis:

Dr. Heriberto Castilla Valdez

Ciudad de México

Diciembre, 2021

"No existen respuestas definitivas, y ello simplemente porque no existen preguntas finales"

- Mario Bunge

Agradecimientos

Quiero agradecer especialmente a mi madre Ana Luisa Rodríguez Camargo, por todo el apoyo incondicional que me ha brindado durante toda mi vida.

Agradezco a toda mi familia, amigos y compañeros de maestría, ya que sin ellos no hubiera podido tener la motivación necesaria para obtener este grado académico.

Agradezco al Dr. Heriberto Castilla Valdez, por haberme aceptado en su equipo de trabajo y por todo lo que aprendí durante el tiempo que trabajamos juntos.

Agradezco al Grupo de Altas Energías, por todas las reuniones de trabajo, en especial a los Maestros Horacio Crotte y Gabriel Ayala por el apoyo invaluable que me brindaron.

Agradezco al Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV) y a todo su personal académico y no académico, por todas las atenciones brindadas durante mi estancia en esa gran institución.

Agradezco a todos mis Profesores del CINVESTAV que contribuyeron a mi formación profesional.

Agradezco al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico necesario para lograr este grado académico.

Resumen

Los decaimientos de partículas que involucran cambios de sabor con corrientes neutras son considerados decaimientos raros debido a las supresiones teóricas presentes en el Modelo Estándar. El decaimiento $B^0 \rightarrow K_s^0 \mu^+ \mu^-$ entra en esta categoría de decaimientos raros y su distribución angular teórica de las partículas, producto del decaimiento, está en términos de un solo parámetro denotado por F_H , en el modelo estándar el valor teórico de F_H está cerca de su frontera permitida. En esta tesis se utilizaron técnicas estadísticas que consideran las restricciones de los parámetros estimados, se obtuvieron intervalos de confianza de F_H por medio de pseudo datos aleatorios generados con la técnica de Toys Monte Carlo, la librería zfit, SLSQP y un método algorítmico proporcionado por el trabajo de Karbach [1] que incluye la técnica de Feldman-Cousins [2]. De los resultados obtenidos se concluyó que las técnicas utilizadas no se obtuvieron sesgos significativos y se logró consistencia del parámetro estimado, por último, los resultados mostraron que la cobertura obtenida de los intervalos de confianza fue aproximadamente válida para una muestra representativa.

Abstract

Particle decays that involve flavor changes with neutral currents are considered rare decays due to the theoretical suppressions present in the Standard Model. The decay $B^0 \rightarrow K_s^0 \mu^+ \mu^-$ falls into this category of rare decays and their theoretical angular distribution of particles, a product of the decay, is in terms of a single parameter denoted by F_H , where its theoretical value is close to its allowed boundary. In this thesis statistical techniques were used that consider the restrictions of estimated parameters, confidence intervals of F_H were obtained by means of pseudo-random data generated from Toys Monte Carlo technique, the zfit library, SLSQP, and an algorithmic method provided by the work of Karbach [1] which includes the Feldman-Cousins technique [2]. From the results obtained, it was concluded that the techniques used did not result in significant biases and consistency of the estimated parameter was achieved. Finally, the results showed that the coverage obtained from the confidence intervals was approximately valid for a representative sample.

Lista de Figuras

1-1.	Modelo estándar de la física de partículas.	2
1-2.	Diagrama de Feynmann que representa un cambio de sabor de quarks por medio de un acoplamiento con la corriente cargada electrodébil W^\pm	5
1-3.	Acoplamiento de quarks q_i , $i = u, c, t, d, s, b$ y bosones neutros. Estos acoplamientos no cambian el sabor de los quarks.	6
1-4.	Diagramas de Feynmann asociados al decaimiento $b \rightarrow sll$. Las partículas dentro del bucle (loop) son partículas virtuales. En la teoría es posible obtener información física del decaimiento exclusivo $B^0 \rightarrow K_s^0 \mu^+ \mu^-$ usando estos diagramas.	8
1-5.	Diagrama esquemático del decaimiento $B^0 \rightarrow K_s^0 \mu^+ \mu^-$	10
3-1.	Ajuste de modelos de probabilidad a datos mediante un optimizador (minimizador).	18
3-2.	Distribución muestral de los estimadores (histograma) y ajuste gaussiano (curva roja). Los valores de μ y σ presentados junto a sus errores son los correspondientes al ajuste Gaussiano.	24
3-3.	Distribución de pulls asociados a diferentes valores del parámetro μ	25
4-1.	Tres posibles opciones de regla de ordenación: intervalo central (izquierda), límite superior (derecha) y límite inferior (abajo), todos a un nivel de confianza $1 - \alpha$	30
4-2.	Cinturones de confianza clásicos de intervalos centrales (izquierda) y límites superiores (derecha). Las líneas rojas horizontales son algunos de los intervalos de aceptación.	30
4-3.	Inversión del cinturón de confianza para la obtención del intervalo de confianza $[\theta_{inf}(x_0), \theta_{sup}(x_0)]$ dado una medición x_0	31
4-4.	Región de confianza S formada por todos los intervalos de confianza que contienen al parámetro verdadero θ_v . La región está delimitada por las dos líneas verticales verdes.	31
4-5.	(Izquierda) Razón de likelihood de Feldman Cousins para una Gaussiana con $\mu = 2$. El valor de k_α (línea horizontal) define el intervalo de aceptación $[x_{inf}, x_{sup}]$. (Derecha) El intervalo de aceptación obtenido define un área en la densidad de probabilidad con un nivel de confianza $1 - \alpha$	33

4-6.	Razón de likelihood para un Gaussiana con $\mu = 0.2$ donde los efectos de la frontera son evidentes.	34
4-7.	Función cumulativa de $\Delta\chi_{toys}^2$ (curva azul) y $\Delta\chi_c^2$ (línea roja vertical). La cantidad $\Delta\chi_c^2$ queda determinada por el nivel de confianza elegido $1 - \alpha$ (línea horizontal verde), en este caso 0.90.	36
4-8.	(Izquierda) Distribución de x_{toy} (histograma), curva $\Delta\chi_{toys}^2$ (curva negra) y $\Delta\chi_c^2$ (recta horizontal roja). (Derecha) Cinturón de confianza resultante.	36
4-9.	Cobertura obtenida para $I_T = 10000$. Los resultados muestran fluctuaciones alrededor de $1 - \alpha = 0.90$	38
4-10.	(Izquierda) Curva $1 - CL$ para una Gaussiana con restricción $\hat{\mu} = \max[0, x_{toy}]$ para la observación $x_{obs} = 1.0$. (Derecha) Curva $1 - CL$ sin restricción. Los intervalos de confianza está dado por la intersección de las rectas horizontales con la curva negra.	40
4-11.	Diagrama de operaciones del algoritmo extendido para un valor de $\mu = \mu_0$	42
4-12.	(Izquierda) Distribuciones $\Delta\chi_{toys}^2$ obtenidas con el algoritmo extendido para todos los valores de μ del intervalo $[1.00, 2.00]$. (Derecha) Distribución de $\Delta\chi_{toys}^2(\mu_0)$ para un valor de $\mu = \mu_0$ y $\Delta\chi_{datos}^2$ con el cual se obtiene la fracción $1 - CL$ tal que $\Delta\chi_{datos}^2 < \Delta\chi_{toys}^2$	42
4-13.	(Izquierda) Curva $1 - CL$ para un conjunto de datos observados con estimador $\hat{\mu}_{obs} = 1.503$ cuyo intervalo de confianza contiene al parámetro verdadero $\mu_v = 1.5$. (Derecha) Curva $1 - CL$ con intervalo de confianza de datos observados que no cuyo intervalo de confianza no contiene al parámetro verdadero.	43
4-14.	Visualización de los intervalos de confianza generados con el algoritmo extendido que cubren o no al parámetro verdadero.	44
5-1.	Estimación del parámetro F_H para $N = 2000$ toy MC junto con sus errores asociados al ajuste Gaussiano.	47
5-2.	Curvas $1 - CL$ e intervalos de confianza (límites) del parámetro F_H que contienen al parámetro verdadero $F_H = 0.2$	48
5-3.	Curvas $1 - CL$ e intervalos de confianza (límites) del parámetro F_H que no contienen al parámetro verdadero $F_H = 0.2$	49
5-4.	Visualización de los intervalos de confianza obtenidos de cada una de las 500 curvas $1 - CL$ generadas para los niveles de confianza de 68% y 90%.	50
5-5.	Visualización de los intervalos de confianza obtenidos de cada una de las 500 curvas $1 - CL$ generadas para los niveles de confianza de 95% y 99%.	51

Lista de Tablas

3-1.	Valores obtenidos de la estimación paramétrica del parámetro μ de una Gaussiana.	23
4-1.	Intervalos de confianza obtenidos al aplicar el algoritmo de Karbach para diferentes valores de x_{obs} a un nivel de confianza de 0.90.	37
4-2.	Comportamiento asintótico de $\beta \rightarrow 1 - \alpha$ (cuadrícula gris) para diferentes valores de μ e I_T	38
4-3.	Intervalos de confianza usando el método de curvas $1 - CL$ de Karbach. Los resultados coinciden con los reportados por Feldman-Cousins.	40
4-4.	Cobertura obtenida usando el algoritmo extendido para distintos valores de μ	43
5-1.	Intervalos de confianza numéricos para unos estimadores $\hat{F}_{H,obs}$ obtenidos a partir del parámetro verdadero $F_H = 0.2$	50
5-2.	Coberturas obtenidas de todos los intervalos de confianza generados.	51

Contenido

Agradecimientos	iii
Lista de Figuras	vi
Lista de Tablas	viii
1. Motivación Teórica	2
1.1. Introducción	2
1.2. Cambios de Sabor con Corrientes Cargadas (CSCC)	5
1.3. Cambios de Sabor con Corrientes Neutras (CSCN)	6
1.4. Distribución Angular y Parámetro F_H	9
2. Probabilidad y Estadística Básica	11
2.1. Probabilidad en Física de Partículas Experimental	11
2.2. Espacio de Probabilidad	12
2.3. Densidad de Probabilidad	12
2.4. Función Likelihood	13
3. Estimación de Parámetros	17
3.1. Ajuste de Modelos	17
3.2. Estimador de Máximo Likelihood	19
3.3. Propiedades de los Estimadores	20
3.4. Varianza de un Estimador	20
3.5. Método Toys Monte Carlo	22
4. Estimación de Intervalos de Confianza	26
4.1. Intervalos de Confianza	26
4.2. Construcción de Neymann	28
4.3. Método de Feldman-Cousins	32
4.4. Método Algorítmico de Karbach	34
4.4.1. Cinturón de Confianza	34
4.4.2. Curvas $1 - CL$	39
4.4.3. Curvas $1 - CL$ Algoritmo Extendido	40

5. Aplicación y Resultados: Decaimiento $B^0 \rightarrow K_s^0 \mu^+ \mu^-$	45
5.1. Estimación del Parámetro F_H	45
5.2. Intervalos de Confianza y Cobertura de F_H	47
5.2.1. Curvas 1-CL para F_H	47
5.3. Conclusiones	52
Apéndice A. zfit y SLSQP	53
Bibliografía	55

1. Motivación Teórica

1.1. Introducción

La física de partículas es una rama de la física que estudia las partículas elementales y sus interacciones. Las partículas elementales son aquellas partículas que no tienen subestructura, en otras palabras, son los objetos que no están hechos de objetos más pequeños [3]. Nuestro actual entendimiento de esta área de la física está englobado en el Modelo Estándar (Figura 1-1). El Modelo Estándar (ME) es una teoría de campos interactuantes que es consistente con las teorías de mecánica cuántica y relatividad especial, en otras palabras es una teoría de campos cuánticos (TCC). El modelo estándar afirma que el material del universo está formado por fermiones elementales que interactúan a través de campos donde las partículas asociadas con los campos de interacción son bosones [4].

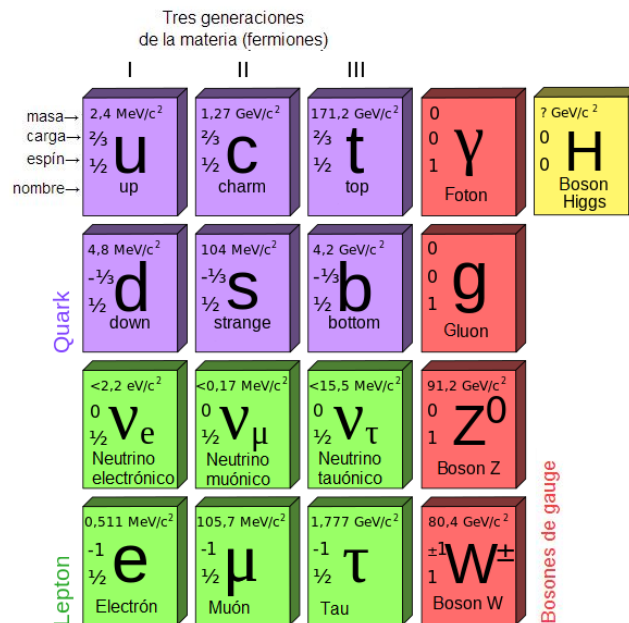


Figura 1-1.: Modelo estándar de la física de partículas.

Cuatro tipos de campos interactuantes se han distinguido en la naturaleza. A la escala de física de partículas, las fuerzas gravitacionales son insignificantes. El modelo estándar excluye al campo gravitacional. Los otros tres campos de interacción están constituidos por los bosones W^+ , W^- y Z^0 , estos bosones son mediadores de la interacción débil, el fotón es el mediador de la interacción electromagnética y el gluón de la interacción nuclear fuerte. Los fermiones elementales del modelo estándar se clasifican en dos tipos: leptones y quarks. Todos los fermiones tienen espín $\frac{1}{2}$ en unidades de \hbar . Los leptones interactúan por medio de la interacción electromagnética y débil, en cambio, los quarks interactúan por medio de la interacción débil, fuerte y electromagnética. Las partículas compuestas de quarks son llamados hadrones, estos se clasifican en bariones (partículas constituidas por 3 quarks) y mesones (partículas constituidas por un quark y un antiquark), por ejemplo el mesón B^0 está constituido por un quark d y un antiquark b .

Desde un punto de vista más formal, la teoría que describe las interacciones de las partículas está estructurada de acuerdo a un grupo de gauge que es invariante bajo las siguientes transformaciones de simetría:

$$SU(3)_{\text{color}} \times SU(2)_{\text{débil}} \times U(1)_{\text{hipercarga}}. \quad (1-1)$$

La simetría $SU(3)_{\text{color}}$ es la simetría de color que describe las interacciones fuertes (cromodinámica cuántica) y $SU(2)_{\text{débil}} \times U(1)_{\text{hipercarga}}$ es la simetría de isospín e hipercarga que describe las interacciones electrodébiles (electromagnética y débil), estos dos grupos de simetría están incluidas en el Lagrangiano del modelo estándar

$$\mathcal{L}_{ME} = \mathcal{L}_{SU(3)} + \mathcal{L}_{SU(2) \times U(1)} \quad (1-2)$$

La física del sabor es un área dentro del modelo estándar. El Lagrangiano electrodébil $\mathcal{L}_{SU(2) \times SU(1)}$ contiene a las interacciones débiles y de Yukawa, estas son las interacciones que estudia la física del sabor [5]. El término sabor es usado para caracterizar diferentes copias de partículas o especies con los mismos números cuánticos de sabor [6], por ejemplo los números cuánticos de sabor de los quarks son:

- Tercera componente de isoespín I_z .
- Extrañeza S .
- Encanto C .
- Número cuántico fondo B .
- Número cuántico cima T .

en otras palabras, la física del sabor es el estudio entre los diferentes sabores de quarks y leptones, su espectro y sus transiciones (cambios de sabor). De la Figura 1-1 se pueden identificar los seis diferentes sabores de quarks: arriba (u), abajo (d), extraño (s), encantado (c), fondo (b) y cima (t) y seis diferentes tipos de sabores leptónicos: electrón (e), muón (μ), tau (τ),

neutrino electrónico (ν_e), neutrino muónico (ν_μ) y neutrino tauónico (ν_τ). Para lo que resta de este capítulo la interacción fuerte no será de interés ya que no ésta no produce cambios de sabor debido a que las interacciones fuertes conservan el sabor.

En esta tesis es de interés el decaimiento dado por $B^0 \rightarrow K_s^0 \mu^+ \mu^-$, decaimiento que involucra un cambio de sabor dado por la transición de quarks $b \rightarrow s$, para comprender mejor esta transición es necesario definir unos conceptos asociados a la teoría electrodébil. En teoría electrodébil el eigenestado débil ψ está compuesto de un doblete izquierdo (I) y dos singletes derechos (D) del grupo de gauge débil $SU(2)$ [7], expresados por

$$\psi = \begin{pmatrix} \psi_I \\ \psi_D \end{pmatrix}, \quad \psi_I = \begin{pmatrix} u_I \\ d_I \end{pmatrix}, \quad \psi_D = u_D, d_D \quad (1-3)$$

donde u_I representa todos los quarks arriba (u, c, t) y d_I representa todos los quarks abajo (d, s, b). Una notación alternativa de (1-3) se puede escribir como

$$\psi_I = \left(\begin{pmatrix} u \\ d \end{pmatrix}_I, \begin{pmatrix} c \\ s \end{pmatrix}_I, \begin{pmatrix} t \\ b \end{pmatrix}_I \right), \quad u_D = \begin{pmatrix} u \\ c \\ t \end{pmatrix}_D, \quad d_D = \begin{pmatrix} d \\ s \\ b \end{pmatrix}_D \quad (1-4)$$

los quarks interactúan con los campos gauge obedeciendo la simetría $SU(2) \times U(1)$, el Lagrangiano de interacción \mathcal{L}_{Int} es dado por

$$\mathcal{L}_{Int} = \bar{\psi}_I (i \not{D}_I) \psi_I + \bar{\psi}_D (i \not{D}_D) \psi_D \quad (1-5)$$

donde $\not{D} = \gamma_\mu D^\mu$ y γ_μ es la matriz de Dirac y las derivadas covariantes D^μ se definen como

$$D_I^\mu = \partial^\mu + ig T^i W_\mu^i + ig' \frac{Y}{2} B_\mu \quad (1-6)$$

$$D_D^\mu = \partial_\mu + ig' \frac{Y}{2} B_\mu \quad (1-7)$$

Las cantidades g y g' son las constantes de acoplamiento de los grupos $SU(2)$ y $U(1)$ respectivamente, W_μ^i , $i = 1, 2, 3$ y B_μ son los campos de gauge (sin masa), T^i es el generador de isoespín débil y Y es el generador de hipercarga. Los bosones electrodébiles físicos (W_μ^\pm , Z_μ , y el fotón A_μ) se expresan como combinaciones lineales de los campos de gauge W_μ^i y B_μ , dado por

$$W_\mu^\pm = \frac{1}{\sqrt{2}} (W_\mu^1 \mp W_\mu^2) \quad (1-8)$$

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix} \quad (1-9)$$

donde θ_W es el ángulo de mezcla débil, los bosones (W_μ^\pm, Z_μ, A_μ) son bosones masivos que adquieren su masa después del rompimiento espontáneo de simetría por medio del mecanismo de Higgs[8]. Los bosones W_μ^\pm participan en las interacciones que cambian el sabor de los fermiones.

1.2. Cambios de Sabor con Corrientes Cargadas (CSCC)

Los bosones cargados electrodébiles W_μ^\pm tienen la propiedad de cambiar el sabor de los quarks y su carga eléctrica por medio de un acoplamiento como se muestra en el diagrama de Feynmann de la Figura 1-2, este cambio de sabor se da entre quarks arriba (u, c, t) con carga $\frac{2}{3}$ y quarks abajo (d, s, b) con carga $-\frac{1}{3}$, estos procesos son conocidos como transiciones con cambios de sabor de corriente cargada.

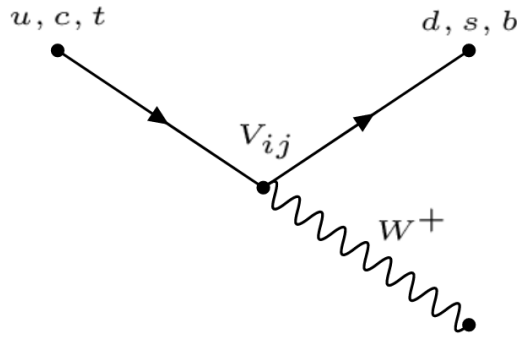


Figura 1-2.: Diagrama de Feynmann que representa un cambio de sabor de quarks por medio de un acoplamiento con la corriente cargada electrodébil W^\pm .

De la Figura 1-2, V_{ij} es el vértice de acoplamiento y corresponde a los elementos de la matriz CKM (Cabibbo, Kobayashi y Maskawa). La matriz CKM se puede derivar partiendo del Lagrangiano de Yukawa [7], sin embargo, en esta tesis solo se presenta su definición. La matriz de CKM es una matriz unitaria 3x3 que describe la probabilidad de transición entre los quarks arriba y quarks abajo, esta matriz corresponde a una rotación entre los eigenestados débiles y los eigenestados de masa de los quarks que surge después del rompimiento de simetría electrodébil, lo que da lugar a mezclas de quarks (cambios de sabor) por medio de transiciones que son proporcional a los factores $|V_{ij}|^2$. Todos los elementos que conforman la matriz CKM está dado por

$$V_{CKM} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix}, \quad V_{CKM} V_{CKM}^\dagger = \bar{1}. \quad (1-10)$$

En el modelo estándar, los acoplamientos entre los bosones neutros (fotón, bosón neutro débil y gluón) y quarks (leptones) no cambian el sabor a nivel árbol, es decir, los diagramas de Feynmann a mínimo orden no tienen como resultado un cambio de sabor como se muestra en la Figura 1-3.

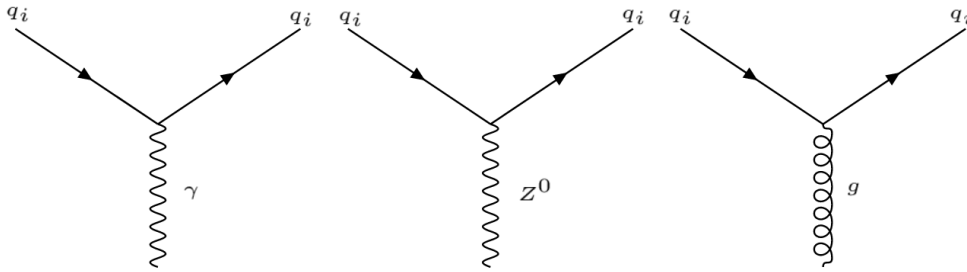


Figura 1-3.: Acoplamiento de quarks q_i , $i = u, c, t, d, s, b$ y bosones neutros. Estos acoplamientos no cambian el sabor de los quarks.

Se puede observar de la matriz CKM que no hay acoplamientos entre la corriente cargada y los quarks de una misma familia, por ejemplo, no hay cambios de sabor a nivel árbol para la transición $b \rightarrow s$, sin embargo, esto no significa que estos procesos de cambio de sabor para esta transición no sean posibles sino que solo no ocurren a nivel árbol [9], las transiciones $b \rightarrow s$ ocurren a órdenes superiores y son transiciones de cambios de sabor con corrientes neutras.

1.3. Cambios de Sabor con Corrientes Neutras (CSCN)

Los cambios de sabor con corrientes neutras (CSCN) son procesos que cambian el sabor de los quarks sin cambiar su carga eléctrica [10]. Un ejemplo de este tipo de procesos es la transición de un quark fondo b en un quark extraño s , ambos pertenecen a la misma familia de quarks abajo con carga $-\frac{1}{3}$.

Desde un tratamiento más formal se puede demostrar que no existen CSCN a nivel árbol [7], si partimos del Lagrangiano de interacción \mathcal{L}_{Int} entre bosones electrodébiles físicos y quarks, este se puede escribir en términos de las corrientes electrodébiles

$$\mathcal{L}_{Int} = -eJ_{EM}^\mu A_\mu - \frac{g}{2}(J_{CC}^\mu W_\mu^+ + J_{CC}^{\mu\dagger} W_\mu^-) - \frac{g}{2\cos\theta_W}J_{CN}^\mu Z_\mu \quad (1-11)$$

estos tres términos contienen las tres corrientes electrodébiles, esto es, la corriente electromagnética J_{EM}^μ , la corriente débil cargada J_{CC}^μ y la corriente débil neutra J_{CN}^μ , como se muestra a continuación:

$$J_{EM}^\mu = \frac{2}{3}\bar{u}\gamma^\mu u - \frac{1}{3}\bar{d}\gamma^\mu d \quad (1-12)$$

$$J_{CC}^\mu = \bar{u}_I\gamma^\mu d_I \quad (1-13)$$

$$J_{CN}^\mu = \bar{u}_I\gamma^\mu \left(1 - \frac{4}{3}\sin^2\theta_W\right) u_I - \bar{u}_D\gamma^\mu \frac{4}{3}\sin^2\theta_W u_D \quad (1-14)$$

$$- \bar{d}_I\gamma^\mu \left(1 - \frac{2}{3}\sin^2\theta_W\right) d_I + \bar{d}_D\gamma^\mu \frac{4}{3}\sin^2\theta_W d_D \quad (1-15)$$

donde $u_{I,D}$ son los quarks arriba y $d_{I,D}$ son los quarks abajo (izquierdos y derechos). Sin entrar en más detalle analíticos, de estas corrientes se puede observar sus acoplamientos fundamentales entre quarks y con esto hacer conclusiones:

Las corrientes cargadas son entre quarks arriba y quarks abajo

- Son interacciones que cambian el sabor
- Ejemplos son las transiciones $t \rightarrow Wb$ y $b \rightarrow Wc$

Los términos de corrientes neutra (y también de corriente electromagnética) solo contienen interacción entre quarks arriba y su respectivo antiquark arriba, lo mismo para los quarks abajo, de esto se concluye que los CSCN

- No cambian el sabor de los quarks.

Estas conclusiones son importantes ya que si se pretende investigar los CSCN se sabe por los resultados anteriores que estas transiciones no ocurren a nivel árbol en el modelo estándar, lo que se traduce transiciones con mayor supresión (debido a la mayor cantidad de vértices presentes en los diagramas de Feynmann a órdenes superiores) y la evidencia experimental obtenida de los colisionadores lo corrobora, de los datos del Particle Data Group las fracciones ramificadas para el decaimiento con CSCC $B^0 \rightarrow D^-\tau^+\nu_\tau$ es de

$$BR(B^0 \rightarrow D^-\tau^+\nu_\tau) = 0.0108 \pm 0.0023 \quad (1-16)$$

mientras que para el decaimiento con CSCN $B^0 \rightarrow K^0\mu^+\mu^-$ está dado por

$$BR(B^0 \rightarrow K^0\mu^+\mu^-) = (3.39 \pm 0.34) \cdot 10^{-7}. \quad (1-17)$$

La diferencia significativa entre las dos fracciones ramificadas muestra que los decaimientos con corriente neutra del meson B^0 decaen con mucho menos frecuencia al canal $K^0\mu^+\mu^-$ relativo al decaimiento con corriente cargada (1-16).

Además de la prohibición de que no existen CSCN a nivel árbol, se debe tomar en cuenta otro mecanismo de supresión dado por los elementos de la matriz CKM relevantes al decaimiento de interés $b \rightarrow sll$, los diagramas de Feynmann de este decaimiento se muestran en la Figura 1-4.

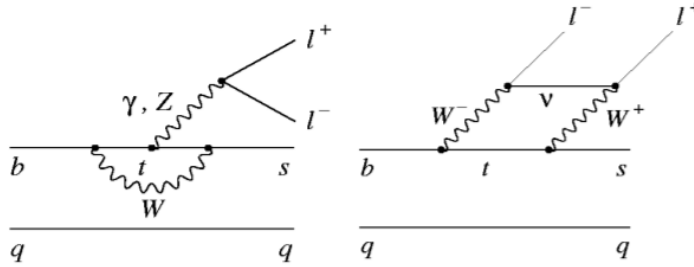


Figura 1-4.: Diagramas de Feynmann asociados al decaimiento $b \rightarrow sll$. Las partículas dentro del bucle (loop) son partículas virtuales. En la teoría es posible obtener información física del decaimiento exclusivo $B^0 \rightarrow K_s^0 \mu^+ \mu^-$ usando estos diagramas.

De la Figura 1-4 se observa que los decaimientos ocurren a nivel de bucle por medio de diagramas de pingüino (figura izquierda) y diagramas de caja (figura derecha), los vértices de CKM relevantes son V_{ts} y V_{tb} , estos vértices entran en los cálculos de la amplitud de decaimiento[11] y ya que V_{ts} es un elemento no diagonal tiene como efecto una supresión significativa a la amplitud de decaimiento. El último mecanismo de supresión está dado por el mecanismo GIM[12].

Por tanto, lo importante de esta sección, es que los cambios de sabor con corriente neutra son procesos que se consideran raros en física de partículas por dos motivos, el primero es por las supresiones teóricas presentes en el modelo estándar (prohibición de nivel árbol, supresión por factor CKM y mecanismo GIM), el segundo son los valores experimentales relativamente pequeños de las fracciones ramificadas de estos decaimientos neutros. Estos resultados tienen como conclusión que debido a la gran supresión en el modelo estándar, los decaimientos con CSCN son lugares ideales para buscar evidencia de nueva física más allá del modelo estándar[13] ya sea por nuevas partículas que sean detectadas por medio de colisiones altamente energéticas o por mediciones más precisas de las propiedades de las partículas que decaen (ej. B^0), tales decaimientos pueden ocurrir a través de partículas transitorias que tienen una masa física mayor que la cantidad de masa-energía disponible de la partícula en descomposición. Estas partículas transitorias se denominan virtuales (ej. quark top y bosón W de la Figura 1-4). Las nuevas partículas pesadas pueden causar grandes desviaciones de las predicciones del modelo estándar. Las mediciones precisas de tales cantidades son sensibles a partículas más allá del modelo estándar que tienen masas que superan con creces la energía de colisión disponible del LHC (Gran Colisionador de Hadrones) [10].

A pesar de la importancia de las conclusiones anteriores para la física de partículas, una conclusión relevante para esta tesis es que los decaimientos de partículas que involucran CSCN son considerados decaimientos raros y por tanto, la información física que se extrae de las mediciones pertinentes pueden tener valores cerca de la frontera física permitida, de aquí se implica que es necesario usar herramientas estadísticas que tomen en cuenta estas consideraciones y la técnica de Feldman-Cousins toma en cuenta estas situaciones donde los resultados pueden estar cerca de la frontera física permitida y con esto proporcionar resultados con una buena cobertura de los errores o intervalos de confianza correspondientes a las mediciones.

1.4. Distribución Angular y Parámetro F_H

La razón de decaimiento Γ que se estudia en física de partículas está dada por dos contribuciones, una cinemática (espacio fase accesible) y otra dinámica (amplitud de decaimiento), dentro de la región cinemática está la distribución angular de decaimiento de las partículas del estado final, esto es, la región física accesible en términos de variables angulares para las partículas producto. Las distribuciones angulares de interés para esta tesis son las asociadas a los siguientes decaimientos exclusivos:

$$B^+ \rightarrow K^+ \mu^+ \mu^-$$

$$B^0 \rightarrow K_s^0 \mu^+ \mu^-$$

La distribución angular del decaimiento $B^+ \rightarrow K^+ \mu^+ \mu^-$ puede ser parametrizada en términos de dos parámetros [14] dado por

$$\frac{1}{\Gamma} \frac{d\Gamma}{d \cos \theta_l} = \frac{3}{4} (1 - F_H) (1 - \cos^2 \theta_l) + \frac{1}{2} F_H + A_{FB} \cos \theta_l \quad (1-18)$$

donde θ_l es el ángulo entre la dirección del muon $\mu^+(\mu^-)$ y el mesón $K^+(K^-)$ a la partícula que decae $B^+(B^-)$ respectivamente (Figura 1-5). Los parámetros de (1-18) son los parámetros A_{FB} y F_H , el primero es la asimetría “adelante-atrás” (forward-backward) del sistema del dimuón y el segundo corresponde a la contribución fraccionaria de las amplitudes escalares y tensoriales a la anchura de decaimiento en la aproximación de muones sin masa. Las predicciones del modelo estándar para estos parámetros son $A_{FB} = 0$ y $F_H \approx 0$ [15] y las constricciones son

$$0 \leq F_H \leq 3 \quad (1-19)$$

$$| A_{FB} | \leq \frac{F_H}{2}. \quad (1-20)$$

Haciendo $A_{FB} = 0$ y multiplicando por un factor de 2 a la ecuación (1-18) se puede eliminar cualquier asimetría residual haciendo análisis en términos de $| \cos \theta_l |$, lo que resulta en la

distribución angular para el decaimiento $B^0 \rightarrow K_s^0 \mu^+ \mu^-$, esta distribución es:

$$\frac{1}{\Gamma} \frac{d\Gamma}{d \cos \theta_l} = \frac{3}{2} (1 - F_H) (1 - |\cos \theta_l|^2) + F_H \quad (1-21)$$

la restricción $0 \leq F_H \leq 3$ prevalece tal que la distribución (1-21) se mantenga positiva para todos los valores de $|\cos \theta_l|$.

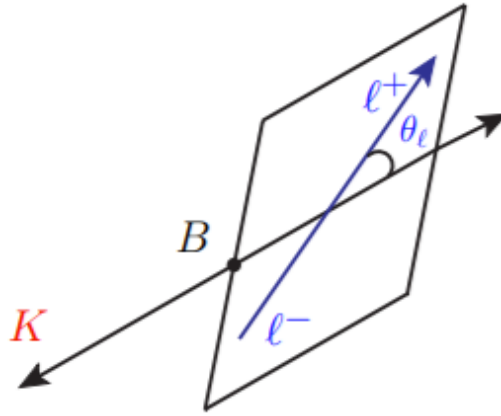


Figura 1-5.: Diagrama esquemático del decaimiento $B^0 \rightarrow K_s^0 \mu^+ \mu^-$.

La distribución angular (1-21) se usará en esta tesis con el objetivo de obtener resultados estadísticos asociados al parámetro F_H usando técnicas de estimación paramétrica y la técnica de Feldman-Cousins para obtener intervalos de confianza y cobertura del parámetro F_H mediante la generación de pseudo datos.

2. Probabilidad y Estadística Básica

2.1. Probabilidad en Física de Partículas Experimental

Dentro de la física de partículas, la teoría de la probabilidad es una herramienta matemática indispensable para extraer información de los sistemas físicos que son investigados, esta teoría es un marco conceptual que evalúa el conocimiento de los procesos aleatorios, sin embargo el enfoque no es único y varía dependiendo de la interpretación de probabilidad[16], estos enfoques son el frecuentista y Bayesiano. En este trabajo se seguirá el enfoque frecuentista de probabilidad que se definirá más adelante. El principal objetivo de un físico experimental es el medir cantidades físicas de interés, posiblemente con la mejor precisión, ejemplos de estas cantidades en física de partículas experimental son las masas de hadrones, el tiempo de vida media de las partículas, los ángulos de decaimiento, los parámetros asociados a la física del sabor, etc. Debido a que estas cantidades son resultados de experimentos aleatorios, es necesario definir unos conceptos concretos de probabilidad, empezando por la definición de probabilidad frecuentista.

Desde el enfoque frecuentista, la probabilidad es la frecuencia con la cual un evento ocurre cuando el número de experimentos tiende a infinito[17], esto es:

$$P(A) = \lim_{n \rightarrow \infty} = \frac{\text{Número de ocurrencias del evento } A}{n} \quad (2-1)$$

donde las cantidades físicas de interés permanecen fijas a un valor, por ejemplo las masas de las partículas elementales. Ejemplos de probabilidad frecuentista en física de partículas son: la probabilidad de que el intervalo [5.3-5.7] GeV contenga la masa del mesón B^0 o que el intervalo [0, 0.2] contenga al parámetro F_H de la distribución angular del decaimiento $B^0 \rightarrow K_s^0 \mu^+ \mu^-$. Estos resultados son aleatorios por diversos factores[18]:

- Límite de resolución de los instrumentos de medición
- Errores de medición aleatorios cuasi-irreducibles (ejemplo: efectos térmicos).
- Teoría no determinista (mecánica cuántica)

2.2. Espacio de Probabilidad

Para modelar un experimento aleatorio es necesario definir la tríada $(\Omega, \mathcal{F}, \mathcal{P})$ que en conjunto es llamado *espacio de probabilidad*. El lenguaje matemático es la teoría de conjuntos que se usará para poder definir otros objetos matemáticos de interés, consideraremos al espacio de probabilidad como estructura básica y otros conceptos como *variable aleatoria* o *densidad de probabilidad* serán definidos con base en $(\Omega, \mathcal{F}, \mathcal{P})$. La definición de estos elementos es la siguiente: El espacio muestral Ω es el conjunto de todos los posibles resultados elementales ω , este conjunto puede ser finito, numerable o infinito, como ejemplo es el tiempo de decaimiento de una partícula inestable, para este caso el espacio muestral se define como $\Omega = \{t \in \mathbb{R} \mid t \geq 0\}$ o para un parámetro físico de interés F_H el espacio muestral es $\Omega = \{F_H \in \mathbb{R} \mid 0 \leq F_H \leq 3\}$. El espacio de eventos \mathcal{F} es definido como un conjunto cuyos elementos representan sub conjuntos del espacio muestral, donde cada elemento A de \mathcal{F} es llamado evento. Del ejemplo anterior el tiempo de decaimiento o el parámetro físico F_H los eventos corresponden intervalos de la recta real[19]. Por último tenemos la medida de probabilidad \mathcal{P} es una función entre conjuntos que asigna una probabilidad entre 0 y 1 a cada elemento del espacio de eventos que satisface los siguientes axiomas también conocidos como axiomas de Kolmogorov [17]:

- $\mathcal{P}(A) \geq 0$ para todo evento A en \mathcal{F}
- $\mathcal{P}(\Omega) = 1$
- Si $A_1, A_2, A_3, \dots \in \mathcal{F}$ son mutuamente excluyentes, entonces

$$\mathcal{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathcal{P}(A_i)$$

Estos tres elementos conforman el espacio de probabilidad.

2.3. Densidad de Probabilidad

Consideremos un experimento cuyas realizaciones es caracterizado por una variable aleatoria continua X . El espacio muestral corresponde a todo el conjunto de posible valores que X puede tomar. La probabilidad de observar un valor en el intervalo infinitesimal $[x, x + dx]$ está dada por la densidad de probabilidad

$$\mathcal{P}(X \in [x + dx]) = f(x)dx \tag{2-2}$$

en la interpretación frecuentista, $f(x)dx$ da la fracción de las veces que x es observado en el intervalo $[x, x + dx]$ en el límite de que el número total de observaciones es infinitamente largo.

La densidad de probabilidad es normalizada tal que la probabilidad total es igual a la unidad

$$\int_S f(x)dx = 1 \quad (2-3)$$

donde la región de integración S se refiere al dominio completo de x , esto es, el espacio muestral completo. También es necesario definir la función de distribución o función acumulativa, sea $(\Omega, \mathcal{F}, \mathcal{P})$ un espacio de probabilidad y $X : \Omega \rightarrow \mathbb{R}$ una variable aleatoria, la función acumulativa de la variable aleatoria X es una función $F : \mathbb{R} \rightarrow [0, 1]$ definida como

$$F(x) = \mathcal{P}[X \leq x] \quad (2-4)$$

esto es, la función acumulativa evaluada en un número x cualquiera es la probabilidad de que la variable aleatoria tome un valor menor o igual a x . Usando esta definición también es posible definir a la densidad de probabilidad como

$$F(x) = \mathcal{P}(X \leq x) = \int_{-\infty}^x f(t)dt \quad (2-5)$$

Valor de Expectación y Varianza

La media o valor de expectación de una variable aleatoria se escribe como $E(X)$. Si observamos N valores de la variable aleatoria X , entonces la media de los N valores será aproximadamente igual a $E(X)$ para largos valores de N . Sea X una variable aleatoria continua con una densidad de probabilidad $f_X(x)$, el valor de expectación de X es

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx \quad (2-6)$$

La varianza de una variable aleatoria X es una medida de cuán dispersa está. La varianza mide que tan lejos están los valores de X de la media

$$V(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2. \quad (2-7)$$

2.4. Función Likelihood

Cuando tratamos con datos que son resultados de mediciones en experimentos y queremos interpretarlos para extraer información física entramos en terreno de la estadística y necesitamos, por tanto, herramientas estadísticas que nos ayuden a interpretar los datos recolectados. En general la estadística es una rama de la matemática que se encarga de la recopilación, análisis, interpretación y presentación de masas de datos numéricos.

Función Likelihood

Cuando una variable aleatoria continua o discreta X se mide n veces (o se extrae una muestra de tamaño n de una población infinita) se obtiene un conjunto de datos $\vec{x} = \{x_0, x_1, \dots, x_n\}$. La variable X se distribuye de acuerdo a una densidad de probabilidad $f_X(x | \vec{\theta})$ donde $\vec{\theta} = \{\theta_1, \theta_2, \dots, \theta_p\}$ son parámetros desconocidos, estos parámetros definen lo que es una densidad de probabilidad paramétrica, por ejemplo una variable aleatoria X sigue una densidad de probabilidad Gaussiana $\mathcal{G}(x | \mu, \sigma)$ donde μ y σ son la media y la desviación estándar o los parámetros de la densidad de probabilidad. El producto de las densidades de probabilidades evaluadas en cada realización de X es la *función likelihood* [20], dado por

$$\mathcal{L}(\vec{\theta} | \vec{x}) = \mathcal{L}(\vec{\theta} | x_0, \dots, x_n) = \prod_{i=1}^n f_X(x_i | \vec{\theta}) \quad (2-8)$$

donde el símbolo $|$ es el símbolo condicional, nos sirve para separar los significados entre la densidad de probabilidad y la función likelihood, en el caso de la densidad de probabilidad $f_X(x | \theta)$ se entiende al objeto como una función que depende de la variable x cuyos parámetros están fijos y la función likelihood $\mathcal{L}(\theta | x)$ una “función” que depende de los parámetros (espacio de p parámetros en general) y la variable x queda fija (espacio de datos). La función likelihood es una función aleatoria, ya que depende de muestras aleatorias de datos, esto es, por ejemplo el conjunto de datos que se obtuvieron del experimento. La utilidad de la función likelihood es transmitir información de cantidades desconocidas. Cantidades desconocidas en problemas estadísticos pueden ser parámetros fijos, de la función likelihood podemos hacer estimaciones de estos parámetros de interés y obtener información sobre la física que subyace al sistema que es estudiado, o en nuestro caso a decaimientos de partículas. Otra utilidad de la función likelihood es obtener intervalos de confianza clásicos e intervalos de confianza dado por el cociente de likelihoods, esto se verá más a detalle en los siguientes capítulos, por ahora basta con su definición.

Función Likelihood Extendida

Si el número datos observados n es también una variable aleatoria que sigue una densidad de probabilidad $P(n | \vec{\theta})$ que también puede depende de p parámetros desconocidos, la función likelihood extendida se define como:

$$\mathcal{L} = P(n | \vec{\theta}) \prod_{i=1}^n f(x_i | \vec{\theta}) \quad (2-9)$$

La función de probabilidad extendida aprovecha el número de eventos registrados n como información extra para determinar los parámetros estimados. Usualmente en física de partículas la distribución $P(n | \vec{\theta})$ es una distribución de Poisson [21].

Razón de Likelihood

Con el objetivo de evaluar un parámetro de interés μ (en general un vector de parámetros $\vec{\mu}$) con respecto a su estimador de máximo likelihood $\hat{\mu}$ dado un conjunto de datos $\{x_0, \dots, x_n\} = \vec{x}$ se define la razón de likelihoods $R(\mu)$ dado por:

$$R(\mu) = \frac{\mathcal{L}(\mu | \vec{x})}{\mathcal{L}(\hat{\mu} | \vec{x})} \quad (2-10)$$

el estimador de máximo likelihood $\hat{\mu}$ es el valor de μ que maximiza la función likelihood, por definición (2-12) está restringido a $0 \leq R(\mu) \leq 1$. Cuando el parámetro μ coincide o se aproxima a $\hat{\mu}$ la razón de likelihood se aproxima a $R(\mu) \approx 1$, de este resultado se debe inferir buena compatibilidad entre los datos (o estimador $\hat{\mu}$) y el valor probado de μ , en cambio cuando existe una gran diferencia numérica entre μ y $\hat{\mu}$ se infiere lo contrario, esto es, una gran incompatibilidad entre los datos y el parámetro μ [22].

En el caso de que la función likelihood dependa de dos vectores de parámetros $\vec{\mu}$ y $\vec{\theta}$ (en una dimensión por simplicidad), conocidos como parámetros de interés (PDI) y parámetros nuisance respectivamente y sólo estemos interesados en los parámetros de interés, definimos la razón likelihood de perfil [23], cuya expresión es:

$$R_p(\mu) = \frac{\mathcal{L}(\mu, \hat{\theta}(\mu) | \vec{x})}{\mathcal{L}(\hat{\mu}, \hat{\theta} | \vec{x})} \quad (2-11)$$

El numerador de (2-13) es la función likelihood de perfil, $\hat{\theta}$ es el valor de θ que maximiza \mathcal{L} para un μ dado, este objeto también es conocido como estimador condicional de máximo likelihood [24], en el denominador de la ecuación (2-13) se tiene a los dos parámetros que maximizan la función likelihood simultáneamente, al igual que (2-12) la razón de likelihood de perfil está restringida al intervalo $[0, 1]$. Aplicando el negativo del logaritmo natural de (2-13) y definiendo la prueba estadística $t_\mu = -2 \ln R_p(\mu)$ en algunos casos es posible obtener la distribución analítica $f(t_\mu | \mu)$ [25] y de esta obtener el p valor que se define como una medida de compatibilidad entre los datos observados (estimador $\hat{\mu}$) y la hipótesis nula μ , entre más grandes sean los valores de p significa una mayor compatibilidad entre los datos observados y la hipótesis nula, esta cantidad se obtiene cuando se compara la prueba estadística t_μ de un conjunto de datos observados $\vec{x}_{obs} \rightarrow t_{\mu,obs}$ con la distribución $f(t_\mu | \mu)$ para un μ de prueba, formalmente el p -valor queda expresado como:

$$p = \int_{t_{\mu,obs}}^{\infty} f(t_\mu | \mu) dt_\mu. \quad (2-12)$$

La función $f(t_\mu | \mu)$ no siempre es posible obtener su forma analítica, sin embargo se puede obtener una aproximación usando métodos computacionales como se verá en el capítulo 4. Las definiciones (2-12)-(2-14) son de suma importancia para esta tesis, ya que se utilizaron fundamentalmente como criterios estadísticos asociados a la técnica de Feldman-Cousin para la

estimación de cinturones de confianza y curvas $1 - CL$ (se definirá más adelante) o curvas de p valor mediante un método algorítmico usando pseudo datos.

3. Estimación de Parámetros

Introducción

En este capítulo se describe como determinar los mejores estimadores de los parámetros asociados a la función likelihood (2-8) dado un conjunto de datos. También se define la distribución muestral de los estimadores para validar las propiedades que estos estimadores deben satisfacer. Por tanto podemos definir la estimación paramétrica como el proceso de determinar y evaluar los estimadores de parámetros desconocidos de un específico modelo probabilístico. Para obtener los estimadores se hace uso del método estadístico de máximo likelihood y se emplea el método de Toys Monte Carlo como fuente de pseudo datos por medio de generación de números pseudo aleatorios.

Las componentes de una estimación paramétrica son las siguientes:

- Un modelo de probabilidad $f(x | \vec{\theta})$, que contiene a los parámetros desconocidos.
- Muestras aleatorias de eventos del modelo de probabilidad.
- Función de costo con la que se obtiene las estimaciones de los parámetros.
- Un conjunto de estimadores de los parámetros desconocidos basados en la información contenida en las muestras aleatorias.
- Las propiedades específicas de los estimadores que se utilizan para evaluar la precisión del estimador.
- Incertidumbres clásicas correspondientes a los estimadores.

El proceso de estimación de parámetros es también llamado ajuste de modelos y es una de las tareas fundamentales de análisis de datos en la física experimental de altas energías [26].

3.1. Ajuste de Modelos

Cuando se ajustan modelos dado por una densidad de probabilidad paramétrica a un conjunto de datos, tenemos que determinar qué tan bien los datos coinciden con el modelo o cuáles son los valores de los parámetros que mejor se ajustan a los datos. Uno de los casos más simples y comunes es cuando se ajustan los datos a una línea recta y se determina la pendiente y la intersección de la recta con el eje de ordenadas, esto es posible extenderlo para casos más

complicados como el que se muestra en la Figura 3-1 por medio de la minimización de una función de costo.

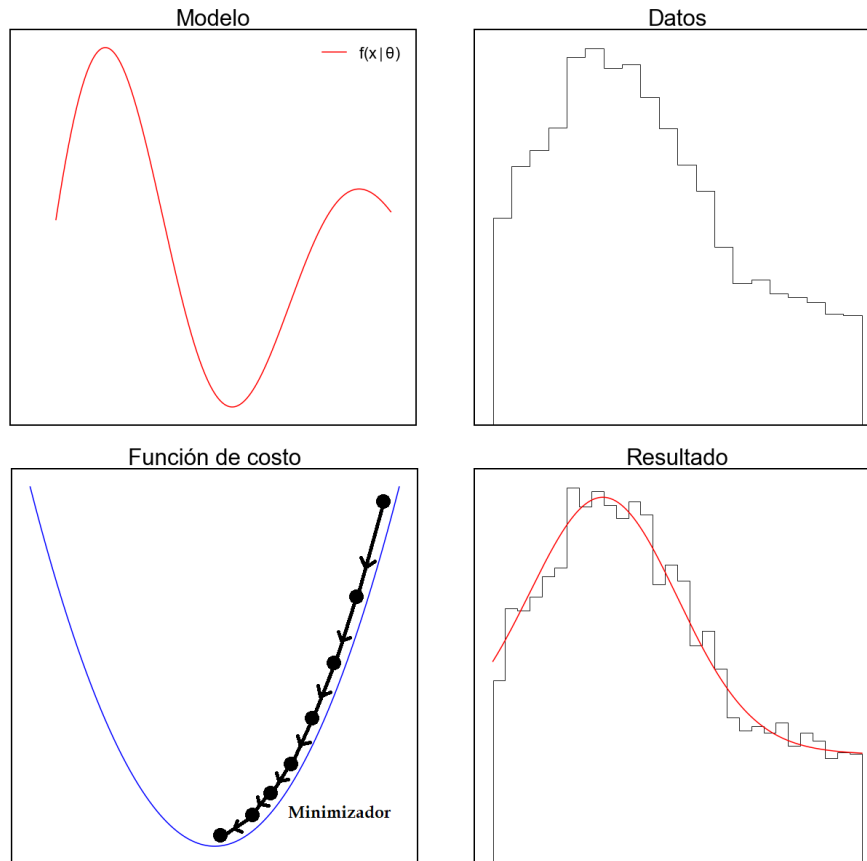


Figura 3-1.: Ajuste de modelos de probabilidad a datos mediante un optimizador (minimizador).

Con el objetivo de encontrar los estimadores de un conjunto de datos se emplea la función likelihood o de “costo” con la cual al maximizar (o minimizar ya sea el caso) mediante un proceso de optimización este método nos proporciona el valor del parámetro que mejor se ajusta a los datos, además de sus errores asociados. Por otro lado, al evaluar el comportamiento de un estimador de un parámetro θ , los valores probables del estimador, la media del estimador y la variabilidad del estimador son características importantes que deben considerarse. El comportamiento de un estimador y su eficacia para estimar un parámetro θ se basan en la *distribución de probabilidad del estimador*, que se denomina distribución muestral. En otras palabras para una muestra aleatoria $\{x_0, \dots, x_n\}$ el estadístico $T = T(x_0, \dots, x_n)$ tiene una distribución de probabilidad f_T asociada a la variable aleatoria T . Se verá más adelante que es posible ajustar una densidad de probabilidad a un conjunto de datos conformado por un vector de estimadores y con esto obtener conclusiones acerca de que tan buenas son las propiedades del estimador.

3.2. Estimador de Máximo Likelihood

Para obtener valores numéricos de los estimadores se emplea el método de máximo likelihood. Con este método se obtiene un conjunto de valores de los parámetros del modelo que maximizan la función likelihood. Intuitivamente, este método maximiza el “acuerdo” entre el modelo y los datos como se ve en la Figura 3.1, siendo la función likelihood la función de costo.

Consideremos la función likelihood dada por la definición (2-10), cuya muestra aleatoria $\{x_0, \dots, x_n\}$ de datos estadísticamente independientes son provenientes de una densidad de probabilidad $f(x | \theta)$ de un solo parámetro θ desconocido, con esto tenemos que la función likelihood es dada por

$$\mathcal{L}(\theta | \vec{x}) = \prod_{i=1}^n f(x_i | \theta) \quad (3-1)$$

Por razones prácticas de cómputo es mejor trabajar con el negativo del logaritmo de la función likelihood. El objetivo es encontrar el valor del parámetro que maximiza la ecuación 3.1 (tanto \mathcal{L} como $\ln \mathcal{L}$ tienen el máximo en el mismo punto), para obtener el parámetro que maximiza \mathcal{L} derivamos (3-1) con respecto al parámetro θ y hacemos la derivada igual a cero

$$\left. \frac{d \ln \mathcal{L}(\theta | \vec{x})}{d\theta} \right|_{\theta=\hat{\theta}} = 0 \quad (3-2)$$

La ecuación (3-2) es se le conoce como ecuación likelihood y el valor de $\hat{\theta}$ que satisface la ecuación es llamado estimador o en particular para este método es llamado estimador de máximo likelihood. Este estimador es una función de los datos $\hat{\theta}(\vec{x})$ y por tanto es también una variable aleatoria. Es posible extender al caso multiparamétrico con p parámetros $\vec{\theta} = \{\theta_1, \theta_2, \dots, \theta_p\}$ con la diferencia de tener que resolver en este caso p ecuaciones likelihood dado por

$$\frac{\partial \ln \mathcal{L}(\vec{\theta} | \vec{x})}{\partial \theta_i} = 0, \quad \forall i = 1, \dots, p. \quad (3-3)$$

Como ejemplo consideremos una densidad de probabilidad Gaussiana dada por la siguiente expresión

$$f(x | \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} \right] \quad (3-4)$$

la densidad de probabilidad (3-4) puede servir para modelar la masa de partículas tanto compuestas como elementales, si la masa es medida n veces cuyas mediciones da un conjunto de datos $\{x_0, \dots, x_n\}$ que se distribuyen normalmente y el objetivo es obtener el estimador del parámetro μ , la función likelihood de (3-4) es

$$\mathcal{L}(\mu, \sigma | \vec{x}) = \prod_{i=0}^n f(x_i | \mu, \sigma) = \prod_{i=0}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_i - \mu)^2}{2\sigma^2}\right] \quad (3-5)$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left[-\frac{\sum_{i=0}^n (x_i - \mu)^2}{2\sigma^2}\right] \quad (3-6)$$

aplicando el logaritmo natural

$$\ln \mathcal{L}(\mu, \sigma | \vec{x}) = n \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{1}{2\sigma^2} \sum_{i=0}^n (x_i - \mu)^2 \quad (3-7)$$

derivando con respecto a μ e igualando a cero

$$\frac{\partial \mathcal{L}(\mu, \sigma | \vec{x})}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=0}^n (x_i - \mu) = 0 \Rightarrow \hat{\mu} = \frac{1}{n} \sum_{i=0}^n x_i \quad (3-8)$$

En este caso particular el estimador máximo likelihood coincide con la media de la muestra.

3.3. Propiedades de los Estimadores

La validez de un estimador queda determinada por sus propiedades. Usando el método de máximo likelihood podemos construir estimadores con las siguientes propiedades:

- Consistencia: un estimador $\hat{\theta}$ es consistente (o asintóticamente consistente) si converge al valor verdadero θ conforme el número de mediciones o datos n se incrementa:

$$\lim_{n \rightarrow \infty} \hat{\theta} = \theta \quad (3-9)$$

- Sesgo: El sesgo (b) es definido como la diferencia entre los valores de expectación del estimador y el parámetro verdadero θ :

$$b = E[\hat{\theta}] - \theta \quad (3-10)$$

El método de máximo likelihood nos permite obtener estimadores consistentes y sin sesgo. Estas tres propiedades se pueden determinar mediante muestras aleatorias de pseudo datos y un ajuste a la distribución muestral de los estimadores.

3.4. Varianza de un Estimador

Ya vimos que podemos estimar el valor de los parámetros desconocidos de una densidad de probabilidad dado un conjunto de datos usando la función likelihood, debido a que el estimador

no es exactamente igual al valor verdadero del parámetro es necesario incluir la incertidumbre asociada al estimador. Los resultados se presentan con la siguiente notación

$$\theta = \hat{\theta} \pm \sigma_{\hat{\theta}} \quad (3-11)$$

donde $\sigma_{\hat{\theta}}$ es la incertidumbre del estimador $\hat{\theta}$. La incertidumbre de un estimador está ligada a su varianza, esta la podemos aproximar desarrollando una serie de Taylor del logaritmo de la función likelihood alrededor del estimador $\hat{\theta}$, esto es

$$\ln \mathcal{L}(\theta) = \ln \mathcal{L}(\hat{\theta}) + (\theta - \hat{\theta}) \left. \frac{d \ln \mathcal{L}(\theta)}{d\theta} \right|_{\theta=\hat{\theta}} + \frac{1}{2} (\theta - \hat{\theta})^2 \left. \frac{d^2 \ln \mathcal{L}(\theta)}{d\theta^2} \right|_{\theta=\hat{\theta}} + \dots \quad (3-12)$$

el segundo término del lado derecho de la ecuación (3-12) es igual a cero por (3-2) y haciendo $-h = \left. \frac{d^2 \ln \mathcal{L}(\theta)}{d\theta^2} \right|_{\theta=\hat{\theta}}$, la función likelihood es proporcional a

$$\mathcal{L}(\theta) \propto \exp \left[-\frac{h}{2} (\theta - \hat{\theta})^2 \right] \quad (3-13)$$

de (3-13) se obtiene la varianza del estimador

$$\sigma_{\hat{\theta}}^2 = h^{-1} = \left(-\left. \frac{d^2 \ln \mathcal{L}(\theta)}{d\theta^2} \right|_{\theta=\hat{\theta}} \right) \quad (3-14)$$

La varianza (3-14) es válida cuando los datos siguen una densidad de probabilidad Gaussiana [26], la incertidumbre o desviación estándar se obtiene tomando la raíz cuadrada de la varianza. Para el caso multiparamétrico, comportamiento Gaussiano y conjuntos de datos finitos la varianza se convierte en la matriz de covarianza \mathbf{V} y se puede aproximar usando la siguiente expresión

$$\mathbf{V} = \left[-\left. \frac{\partial^2 \ln \mathcal{L}(\vec{x} | \vec{\theta})}{\partial \vec{\theta}^2} \right|_{\vec{\theta}=\hat{\theta}} \right]^{-1} = \mathbf{H}^{-1} \quad (3-15)$$

donde \mathbf{H} es la matriz Hessiana y la incertidumbre $\hat{\theta}_i$ se obtiene tomando la raíz cuadrada de los elementos diagonales de (3-15) dado por

$$\sigma_{\hat{\theta}_i} = \sqrt{V_{ii}} \quad (3-16)$$

los intervalos de confianza clásicos se obtienen de las incertidumbres estándar $\sigma_{\hat{\theta}}$ y se pueden escribir como

$$[\hat{\theta} - \sigma_{\hat{\theta}} < \theta < \hat{\theta} + \sigma_{\hat{\theta}}] \quad (3-17)$$

estos son intervalos de confianza simétricos, sin embargo, es posible obtener intervalos de confianza asimétricos (capítulo 4) sin las restricciones mencionadas en esta sección.

3.5. Método Toys Monte Carlo

Con el objetivo de comprobar que los estimadores obtenidos de una densidad de probabilidad cumple las propiedades mencionadas anteriormente se utiliza la técnica de Toys Monte Carlo, que consiste en obtener muestras aleatorias de pseudo datos por medio de un generador algorítmico de números aleatorios que siguen una densidad de probabilidad fija, esto es, la generación de números aleatorios es la obtención sucesiva de la variable aleatoria X de una densidad de probabilidad definida f_X , estos números aleatorios se usan para simular la física del experimento.

Usando el método Monte Carlo es posible obtener la distribución de los estimadores o distribución muestral. Como primera prueba consideramos una densidad de probabilidad Gaussiana $\mathcal{G}(x | \mu, \sigma)$ donde la gaussiana puede modelar una señal asociada a la distribución de masas de una partícula cuyo parámetro de interés es el parámetro μ , procedemos a obtener muestras aleatorias o Toys Monte Carlo de la densidad \mathcal{G} , después de haber obtenido los datos de n eventos se maximiza el likelihood para cada vector de pseudo datos $\{x_0, \dots, x_n\}$ generados y se grafica la distribución del estimador; con más claridad los pasos son los siguientes:

- Se elige una densidad de probabilidad paramétrica (ej. Gaussiana)

$$\mathcal{G}(x | \mu, \sigma) \tag{3-18}$$

- Generar N muestras aleatorias o Toy MC con n eventos

$$\mathcal{G}(x | \mu, \sigma) \rightarrow [\{x_0, \dots, x_n\}_0, \dots, \{x_0, \dots, x_n\}_N]$$

- Aplicar el método de máximo likelihood para cada i -ésimo vector de datos $\{x_0, \dots, x_n\}_i$ extrayendo el estimador $\hat{\mu}$ y su incertidumbre asociada $\sigma_{\hat{\mu}}$ en cada iteración

$$\{x_0, \dots, x_n\}_i \rightarrow \mathcal{L}_i \rightarrow (\hat{\mu}, \sigma_{\hat{\mu}})_i$$

- Se grafica en un histograma del conjunto de estimadores obtenidos en el paso anterior

$$[\hat{\mu}_0, \dots, \hat{\mu}_N] \rightarrow \text{Histograma}$$

- Por último se ajusta la distribución de estimadores a un modelo Gaussiano para obtener el sesgo y consistencia del estimador.

Usando herramientas computacionales, en este caso Python, la librería zfit [27] y Sci-Py, con estas librerías se obtienen las muestras aleatorias de datos y se calculan los estimadores así como las incertidumbres asociadas a cada estimador, con esto podemos obtener resultados del algoritmo anterior y presentarlos tanto numérica como gráficamente. Empleando el algoritmo con los parámetros verdaderos $\mu_v = [1.0, 2.0, 3.0, 5.0]$, $\sigma_v = 1.0$, $n = 1000$ eventos y $N = 500$

muestras (vectores de datos) de los n eventos, se obtiene las distribuciones de los estimadores $\hat{\mu}$ (Figura 3-2). Los resultados que se reportan son los siguientes

μ_v	$E[\hat{\mu}]$	$\sigma_{\hat{\mu}}$	$ b $	$R_d = b / \epsilon_d$
1.0	0.9997 ± 0.0010	0.0316 ± 0.0007	0.0003	0.3000σ
2.0	1.9987 ± 0.0010	0.0305 ± 0.0007	0.0013	1.3000σ
3.0	3.0007 ± 0.0010	0.0326 ± 0.0007	0.0007	0.7000σ
4.0	4.0001 ± 0.0010	0.0317 ± 0.0007	0.0001	0.1000σ

Tabla 3-1.: Valores obtenidos de la estimación paramétrica del parámetro μ de una Gaussiana.

donde ϵ_d es el error obtenido del ajuste a la distribución de $\hat{\mu}$. De los datos obtenidos presentados en la Tabla 3-1 se concluye que el estimador es consistente y con un sesgo despreciable.

Pulls

Para comprobar que los estimadores son consistentes y con poco sesgo se usa también el método de pulls. Si una variable aleatoria X es generada repetidamente de una distribución Gaussiana con media μ y desviación estándar σ , entonces el pull de la distribución se define como

$$\text{Pull} = \frac{X - \mu}{\sigma} \quad (3-19)$$

en nuestro caso anterior X son los estimadores $\hat{\mu}$, μ es nuestro valor verdadero del parámetro y σ es la incertidumbre de $\hat{\mu}$, esto es

$$\text{Pull} = \frac{\hat{\mu} - \mu_v}{\sigma_{\hat{\mu}}} \quad (3-20)$$

Si la distribución de pulls siguen una distribución Gaussiana estándar, es decir una Gaussiana con media cero y desviación estándar igual aproximadamente a la unidad se puede concluir que no hay sesgo [28]. Siguiendo el ejemplo de la Gaussiana los pulls asociados a los resultados de la Figura 3-2 se muestran en la Figura 3-3.

Los valores reportados son los siguientes:

$$R_d(\mu_v = 1.0) = \frac{\mu_p}{\epsilon_p} = \frac{0.0001}{0.0316} = 0.0032 \quad (3-21)$$

$$R_d(\mu_v = 2.0) = 1.1401 \quad (3-22)$$

$$R_d(\mu_v = 3.0) = 0.7025 \quad (3-23)$$

$$R_d(\mu_v = 4.0) = 0.1041 \quad (3-24)$$

donde ϵ_p es el error de ajuste de la distribución de pulls asociados al parámetro μ .

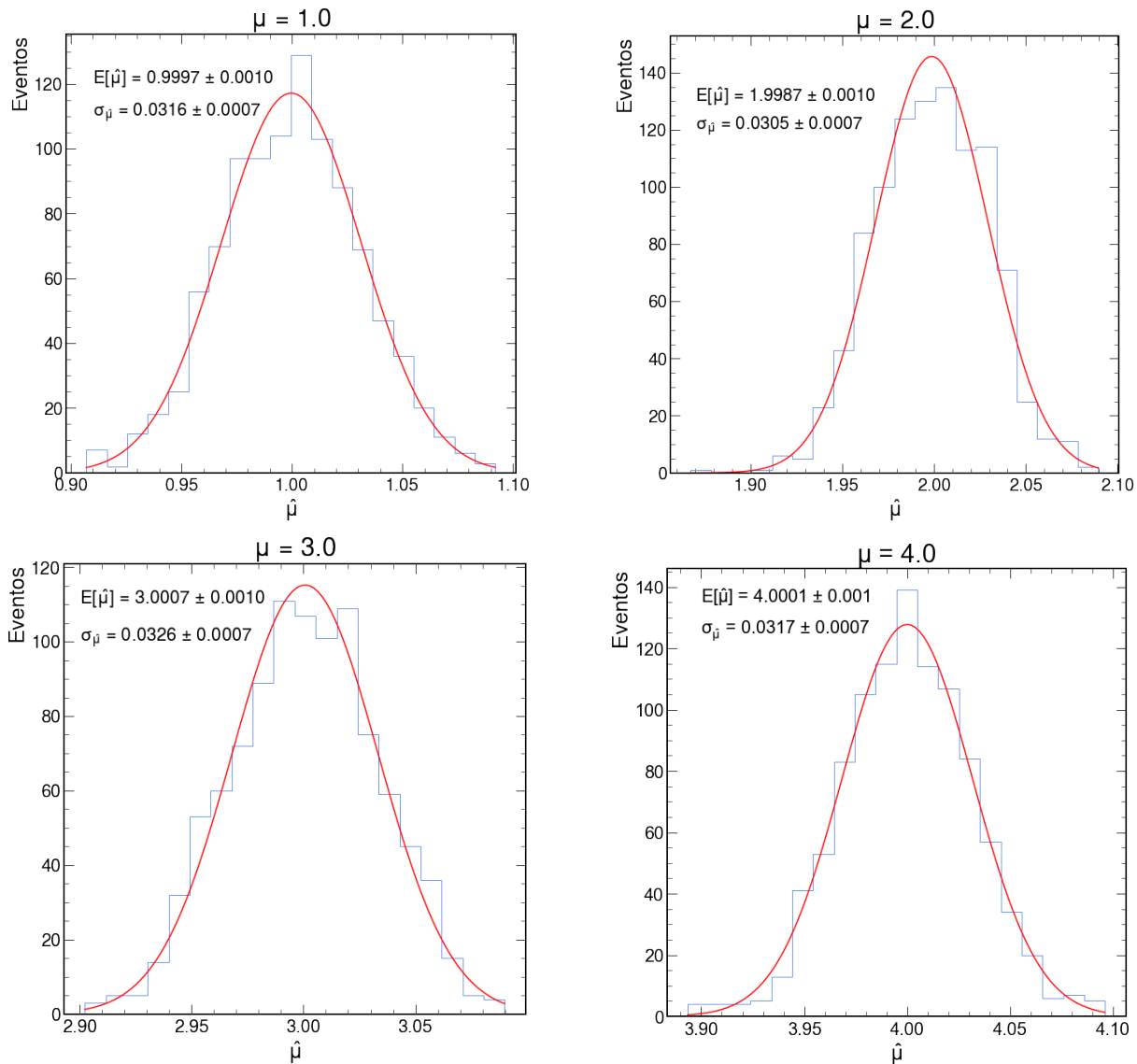


Figura 3-2.: Distribución muestral de los estimadores (histograma) y ajuste gaussiano (curva roja). Los valores de μ y σ presentados junto a sus errores son los correspondientes al ajuste Gaussiano.

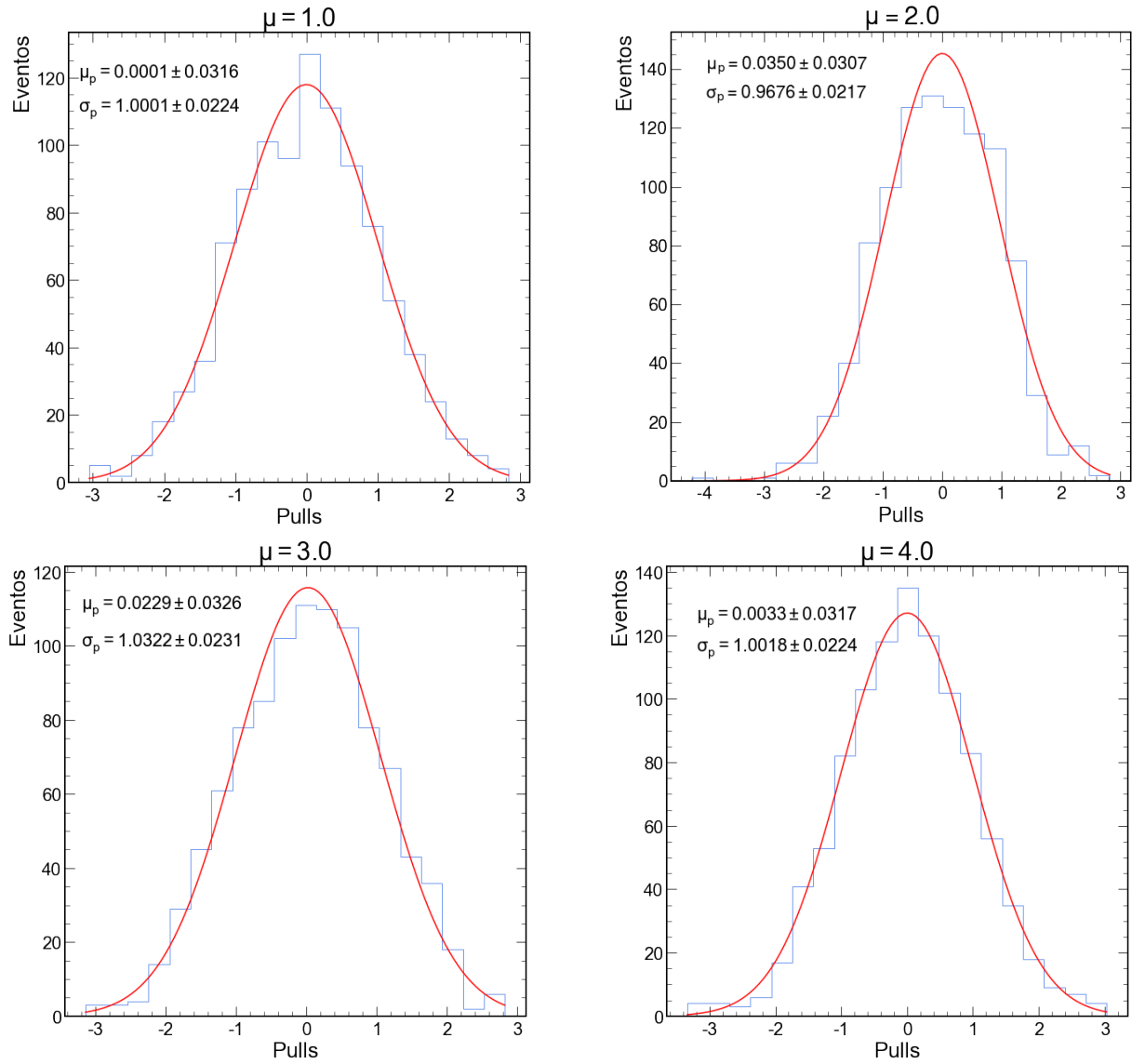


Figura 3-3.: Distribución de pulls asociados a diferentes valores del parámetro μ .

Cada uno de los resultados dado por (3-21)-(3-24) es menor que 3σ , por tanto se concluye que no hay sesgo con nuestra distribución de pulls.

4. Estimación de Intervalos de Confianza

Introducción

La estimación de parámetros descrita en el capítulo 3 nos proporciona información sobre como extraer parámetros con buenas propiedades utilizando el método de máximo likelihood. Los valores numéricos de los estimadores no es lo único que podemos extraer del likelihood de un modelo o de una densidad de probabilidad dado un conjunto de datos, también es posible obtener sus intervalos de confianza, esto es necesario ya que necesitamos una caracterización de la confiabilidad de la estimación, además de considerar las restricciones físicas a las que está impuesto nuestro modelo, la estimación de intervalos cubre estas dos tareas. En lugar de una sola estimación numérica del estimador se proporcionarán dos límites numéricos (simétricos o asimétricos) más un nivel de confianza sobre el valor real del parámetro de interés. Con el objetivo de lograr esta confiabilidad en los intervalos de confianza se siguen dos algoritmos creados por Karbach[1], que en resumen, son métodos algorítmicos que se desarrollan usando pseudo datos por medio del método de Toys Monte Carlo para calcular intervalos de confianza, en estos algoritmos se incluye la regla de ordenación de Feldman Cousins[2]. Por último, los resultados de los intervalos de confianza no son únicos, esto es, son cantidades que dependen de los datos y por tanto es necesario reportar cobertura. En este capítulo se aplicó sistemáticamente tres algoritmos (dos algoritmos de Karbach y un algoritmo adicional) y se obtuvo la cobertura de cada uno para el caso de una densidad de probabilidad Gaussiana, en el capítulo siguiente se aplicaron estos algoritmos al parámetro F_H del decaimiento $B^0 \rightarrow K_s^0 \mu^+ \mu^-$.

4.1. Intervalos de Confianza

En el caso clásico, los resultados que se reportan de un experimento es una estimación de algún parámetro junto con su desviación estándar donde esta última es una medida de que tanto los estimadores pueden distribuirse si el experimento es repetido muchas veces. Como tal, la desviación estándar σ informa la incertidumbre estadística de una medición y se denomina error estándar, este enfoque se presentó en el ejemplo de la Gaussiana del capítulo anterior. Los intervalos de confianza no tienen que seguir esta formulación clásica de incertidumbre asociada a los parámetros de interés, para desarrollar un tratamiento más general primero definimos lo que es un intervalo de confianza:

Definición (Intervalos de confianza): Sea $\{x_0, \dots, x_n\}$ una muestra de la densidad de probabilidad $f(x | \theta)$ donde θ es el parámetro de interés. Supongamos que es posible construir dos funciones de la muestra de valores $\theta_{inf}(x_0, \dots, x_n)$ y $\theta_{sup}(x_0, \dots, x_n)$ de modo que la probabilidad del intervalo aleatorio $[\theta_{inf}, \theta_{sup}]$ cubre al parámetro verdadero θ con una fracción $1 - \alpha$. Esto es

$$\mathcal{P}[\theta_{inf}(x_0, \dots, x_n) \leq \theta \leq \theta_{sup}(x_0, \dots, x_n)] = 1 - \alpha \quad (4-1)$$

La cantidad $1 - \alpha$ es el nivel de confianza, θ_{inf} es el límite de confianza inferior y θ_{sup} es el límite de confianza superior. La longitud $\theta_{sup} - \theta_{inf}$ es el intervalo de confianza.

Para validar la confiabilidad del nivel de confianza y por lo tanto del intervalo de confianza se define la probabilidad de cobertura:

Definición (Probabilidad de cobertura): Dado un conjunto de intervalos de confianza $C([\theta_{inf}, \theta_{sup}])$, la probabilidad de cobertura β es la fracción del conjunto de intervalos de confianza que contienen al parámetro verdadero θ_v

$$\mathcal{P}(\theta_v \in C([\theta_{inf}, \theta_{sup}])) = \beta \quad (4-2)$$

Dado un conjunto finito de experimentos que se realizan bajo condiciones similares y que miden al parámetro verdadero θ_v , la probabilidad de cobertura finita es la fracción de veces que el intervalo de confianza contiene al parámetro verdadero, esto es

$$\mathcal{P}(\theta_v \in C([\theta_{inf}, \theta_{sup}])) \approx \frac{N(C([\theta_{inf}, \theta_{sup}])_V)}{N(C([\theta_{inf}, \theta_{sup}])_T)} = \frac{I_V}{I_T} \quad (4-3)$$

donde $N(C([\theta_{inf}, \theta_{sup}])_V) = I_V$ es el número de intervalos de confianza que contienen al parámetro verdadero y $N(C([\theta_{inf}, \theta_{sup}])_T) = I_T$ es el número total de intervalos de confianza capturados. Si el número total de intervalos de confianza tiende a infinito (enfoque frecuentista) y la probabilidad de cobertura β tiende al nivel de confianza $1 - \alpha$ entonces los intervalos de confianza quedan validados. Un ejemplo es necesario para resaltar el significado de los intervalos de confianza: podemos pensar que la probabilidad de que la masa del mesón $m_B \in [m_1, m_2]$ sea del 90%, sin embargo este enunciado es incorrecto, ya que la masa del mesón B está fija a un valor único, por otro lado si corremos un conjunto de experimentos toy MC, cada experimento tendrá como resultado un intervalo de confianza distinto (característica de un estimador), entonces, la afirmación correcta es: si hay un mesón B con una masa m_B , entonces, en un conjunto de experimentos, el 90% de los intervalos de confianza obtenidos contendrá el valor verdadero de m_B . Esto es parte del enfoque frecuentista que considera que los parámetros verdaderos o “parámetros de la naturaleza” como cantidades fijas y los intervalos de confianza como estimadores (variables aleatorias) cuyos resultados varían en

función de los datos obtenidos del experimento.

Existen propiedades deseables que se deben considerar al construir intervalos de confianza[26]:

- Efecto de las fronteras físicas: Cuando el parámetro de interés tiene fronteras impuestas por alguna restricción, algunas construcciones de intervalos pueden caer en región donde no es físicamente posible interpretar los resultados, un ejemplo son la restricción a las masas de las partículas donde se define positiva y cualquier resultado negativo no es deseable. Por tanto siempre hay que considerar las restricciones físicas impuestas al sistema para obtener intervalos de confianza válidos, por ejemplo la restricción física de masas $\mu \geq 0$ o que el parámetro F_H esté definido en el intervalo $[0 \leq F_H \leq 3]$.
- Relación con la estimación puntual: Cuando se mide la propiedad de un sistema que se sabe que existe usualmente se reportan tanto la estimación puntual como los intervalos de confianza y es deseable que el primero sea incluido en el segundo.
- Validación de los intervalos de confianza: Si la probabilidad de cobertura β es aproximadamente igual al nivel de confianza $1 - \alpha$ entonces los intervalos de confianza son válidos.

No existe un solo método para la obtención de estos intervalos de confianza, sin embargo, en esta tesis usaremos la construcción frecuentista de Neymann y Feldman-Cousins por medio de un algoritmo computacional de Karbach. Se obtendrá la cobertura para validar los intervalos de confianza.

4.2. Construcción de Neymann

En el capítulo 3 vimos como obtener intervalos de confianza clásicos al asumir un likelihood Gaussiano cuyos resultados se presentan como

$$[\hat{\theta} \pm \sigma_{\hat{\theta}}] \quad (4-4)$$

Una construcción más general de intervalos de confianza que sigue el enfoque frecuentista es gracias al trabajo de Neymann. Consideremos una variable aleatoria X que sigue una densidad de probabilidad paramétrica y que depende (por simplicidad) de un sólo parámetro θ . El procedimiento de Neymann para determinar los intervalos de confianza sigue tres pasos:

- Elección de la regla de ordenación.
- Construcción del cinturón de confianza.
- Inversión del cinturón de confianza para determinar el intervalo de confianza.

Primero hay que escoger un conjunto de valores del parámetro de interés

$$[\theta_0, \dots, \theta_\eta] \quad (4-5)$$

donde η es el número total de nuestro conjunto de parámetros. Para cada parámetro de (4-5) se obtiene su densidad de probabilidad correspondiente y con esto obtenemos un conjunto de densidades de probabilidad

$$[\theta_0, \theta_\eta] \rightarrow [f(x | \theta_0), \dots, f(x | \theta_\eta)] \quad (4-6)$$

para cada valor del parámetro $\theta = \theta_\eta$ la densidad $f(x | \theta_\eta)$ es conocida. Ya definido nuestras densidades de probabilidad procedemos a obtener los “intervalos de aceptación” $[x_{inf}(\theta), x_{sup}(\theta)]$, estos intervalos se obtienen tal que cubran con un nivel de confianza deseado $1 - \alpha$ para cada una de las densidades de probabilidad, esto es, la construcción de Neymann de los intervalos de aceptación es tal que la probabilidad

$$P(x \in [x_{inf}, x_{sup}] | \theta) = 1 - \alpha \quad (4-7)$$

Los niveles de confianza que frecuentemente se reportan son niveles de confianza en términos de σ , por ejemplo para el caso unidimensional: $1\sigma = 68.27\%$, $2\sigma = 95.5\%$, $3\sigma = 99.73\%$, etc. La relación 4-7 es equivalente a

$$\int_{x_{inf}}^{x_{sup}} f(x | \theta) dx = 1 - \alpha \quad (4-8)$$

La elección del intervalo de aceptación es arbitraria y se le conoce como regla de ordenación, esta es una regla con la que se decide cuáles valores de x son incluidos en el intervalo de aceptación, la única restricción es que el intervalo debe cumplir con la ecuación (4-8). Las 3 ordenaciones más comunes se muestran en la Figura 4-1.

En esta elección se tiene total libertad siempre y cuando la decisión no esté influenciada por los datos. La probabilidad que corresponde a intervalos superiores es dado por

$$P(x < x_1 | \theta) = 1 - \alpha \quad (4-9)$$

y para intervalos centrales, tenemos

$$P(x < x_1 | \theta) = P(x > x_2 | \theta) = \frac{\alpha}{2} \quad (4-10)$$

Se concluye la construcción cuando se obtienen todos los intervalos de aceptación para cada valor del parámetro θ del intervalo (4-5). El conjunto de todos los intervalos de aceptación definen al cinturón de confianza como se muestra en la Figura 4-2.

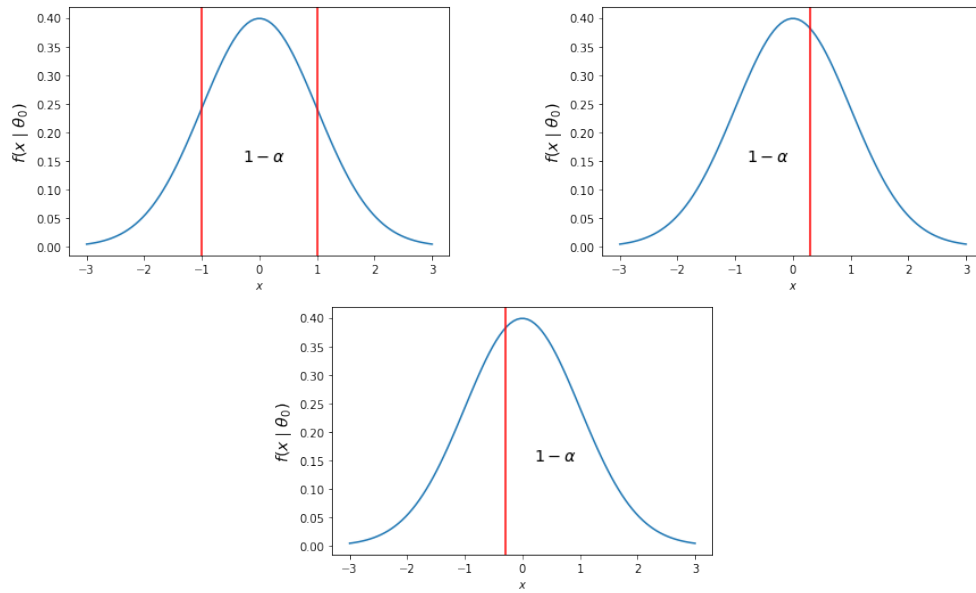


Figura 4-1.: Tres posibles opciones de regla de ordenación: intervalo central (izquierda), límite superior (derecha) y límite inferior (abajo), todos a un nivel de confianza $1 - \alpha$.

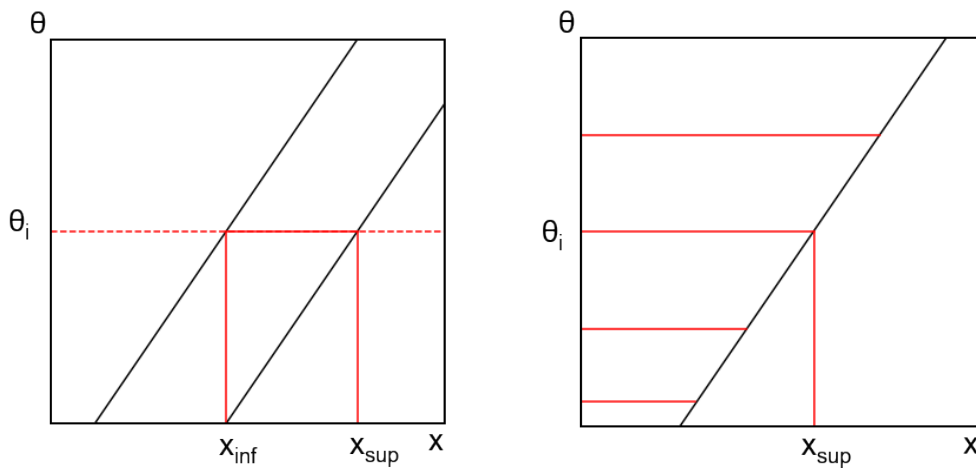


Figura 4-2.: Cinturones de confianza clásicos de intervalos centrales (izquierda) y límites superiores (derecha). Las líneas rojas horizontales son algunos de los intervalos de aceptación.

Lo siguiente es hacer una inversión del cinturón confianza, esto es, al trazar una línea vertical dado por una medición x_0 , ésta se interseca con el cinturón de confianza y las intersecciones dan como resultado un intervalo de confianza $[\theta_{inf}(x_0), \theta_{sup}(x_0)]$ a un nivel de confianza definido como se muestra en la Figura 4-3.

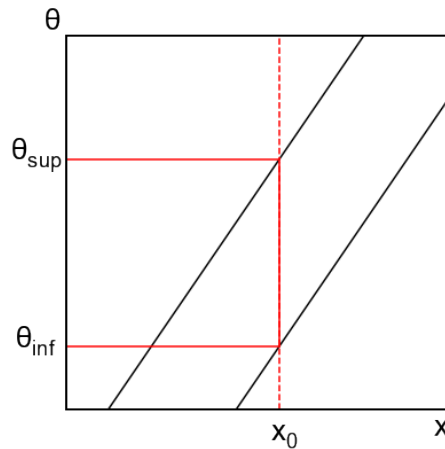


Figura 4-3.: Inversión del cinturón de confianza para la obtención del intervalo de confianza $[\theta_{inf}(x_0), \theta_{sup}(x_0)]$ dado una medición x_0 .

El intervalo $[\theta_{inf}(x_0), \theta_{sup}(x_0)]$ tiene, por construcción, una cobertura igual al nivel de confianza $1 - \alpha$. La obtención de la cobertura asociada al cinturón de confianza es la siguiente: supongamos un valor definido del parámetro verdadero $\theta = \theta_v$, si se hacen N observaciones x_0 de cada una de estas observaciones se obtiene su intervalo de confianza, la fracción de intervalos que contiene a θ_v sobre el total es la probabilidad de cobertura (4-13). La región de confianza S , esto es, la región con los intervalos de confianza que contienen a parámetro verdadero θ_v queda definida por la línea horizontal como se muestra en la Figura 4-4.

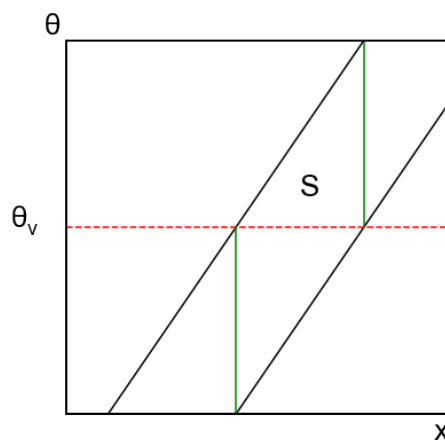


Figura 4-4.: Región de confianza S formada por todos los intervalos de confianza que contienen al parámetro verdadero θ_v . La región está delimitada por las dos líneas verticales verdes.

Esta construcción de Neymann se puede extender para el caso donde el valor de x es igual a los

estimadores [21] $x = \hat{\theta}(x_0, \dots, x_n)$, la construcción es similar solo que en lugar de la densidad de probabilidad $f(x | \theta)$ es sustituida por la densidad de probabilidad del estimador $g(\hat{\theta}(\vec{x}), \theta)$ [17], por tanto el cinturón de confianza obtenido estaría en el plano $(\theta, \hat{\theta})$,

4.3. Método de Feldman-Cousins

Los cinturones de confianza clásicos de Neymann de la sección anterior nos proporcionan intervalos de confianza sin embargo hay valores de x que no intersectan con el cinturón (Figura 4-3), es decir, no se obtienen intersecciones de la recta vertical con el cinturón de confianza y por tanto el intervalo de confianza es igual al vacío, con el objetivo de remediar este problema se empezó a usar la política “flip-flop”, que consiste en dividir en regiones distintas al cinturón de confianza, la primera y segunda región permiten obtener intervalos de límites superiores y la tercera región permite obtener intervalos centrales; esta política produce una cobertura incorrecta. Para obtener cinturones de confianza que preserven cobertura y hagan una transición natural de límites superiores a intervalos centrales, Feldman y Cousins introdujeron una nueva regla de ordenación dado por la razón de de likelihoods (2-10), esto es, para una densidad de probabilidad $f(x | \theta)$ con un parámetro de interés θ la razón de likelihoods que se usa como ordenación está dado por

$$R = \frac{\mathcal{L}(\theta | x)}{\mathcal{L}(\hat{\theta} | x)} > k_\alpha \quad (4-11)$$

donde $\hat{\theta}$ es el valor del parámetro que maximiza la función likelihood y la constante k_α determina el intervalo de aceptación C_α tal que la integral de $f(x | \theta)$ sea igual al nivel de confianza $1 - \alpha$, esto es

$$\int_{C_\alpha} f(x | \theta) dx = 1 - \alpha \quad (4-12)$$

el intervalo de aceptación es dado por todos los valores de x tal que

$$C_\alpha = \{x | R > k_\alpha\}, \quad R(x_{inf}) = R(x_{sup}) \quad (4-13)$$

Para ilustrar este método consideremos una distribución Gaussiana $G(x | \mu)$ con un σ conocido, considerando estimadores físicamente accesibles, esto es, el valor de μ que maximiza el likelihood se obtiene por medio de la siguiente relación

$$\hat{\mu} = \max[0, x] \quad (4-14)$$

la expresión anterior restringe nuestros resultados a valores positivos del estimador $\hat{\mu}$, esta restricción hace que el likelihood Gaussiano se divida en dos regiones acordes al valor de x ,

esto es

$$\mathcal{L}_G(\hat{\mu} | x) = \begin{cases} \frac{1}{\sqrt{2\pi}} & \text{si } x \geq 0 \\ \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) & \text{si } x < 0 \end{cases} \quad (4-15)$$

con esto, nuestra razón de likelihoods queda determinada por

$$R = \frac{\mathcal{L}_G(\mu | x)}{\mathcal{L}_G(\hat{\mu} | x)} = \begin{cases} \exp(-(x - \mu)^2/2) & \text{si } x \geq 0 \\ \exp(x\mu - \mu^2/2) & \text{si } x < 0 \end{cases} \quad (4-16)$$

El intervalo de aceptación $[x_{inf}, x_{sup}]$ se puede obtener numéricamente (sección 4.4) a partir de la ecuación (4-16) por medio de la implementación de un algoritmo y con esto construir el cinturón de confianza de Feldman Cousins. Los valores de x se agregan a la región de aceptación para un μ dado en orden decreciente de R hasta que se alcanza el nivel de confianza deseado (Figura 4-5 y Figura 4-6).

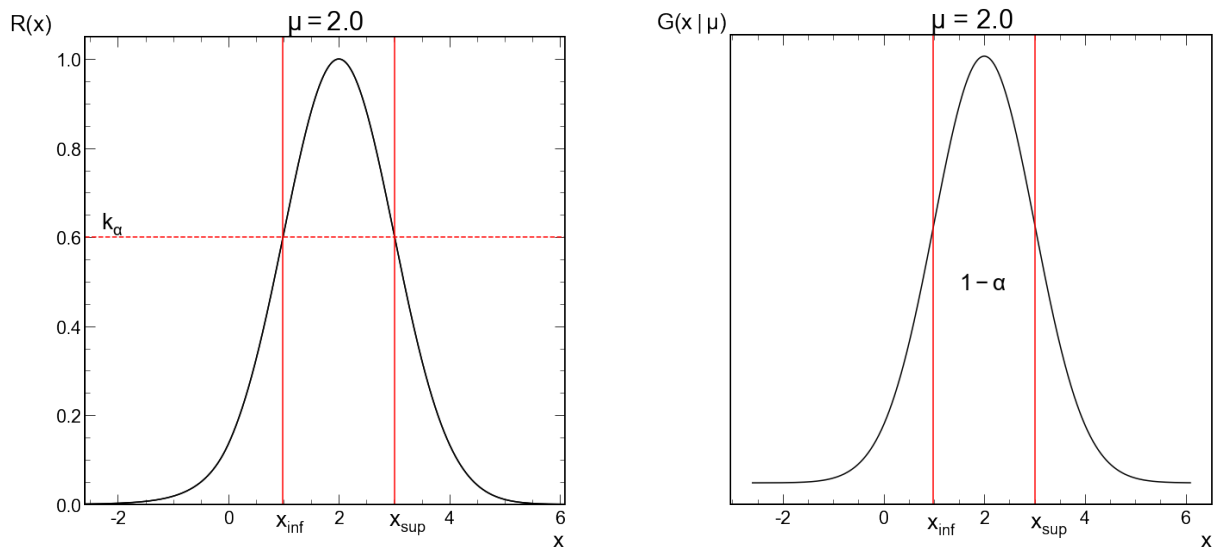


Figura 4-5.: (Izquierda) Razón de likelihood de Feldman Cousins para una Gaussiana con $\mu = 2$. El valor de k_α (línea horizontal) define el intervalo de aceptación $[x_{inf}, x_{sup}]$. (Derecha) El intervalo de aceptación obtenido define un área en la densidad de probabilidad con un nivel de confianza $1 - \alpha$.

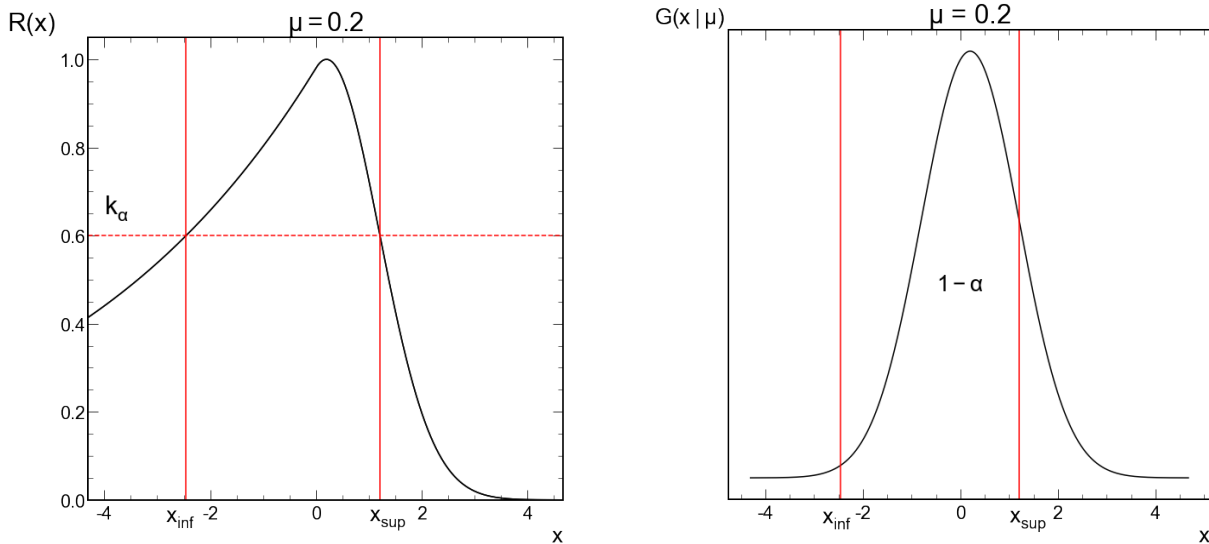


Figura 4-6.: Razón de likelihood para un Gaussiana con $\mu = 0.2$ donde los efectos de la frontera son evidentes.

4.4. Método Algorítmico de Karbach

El método algorítmico de Karbach se divide en dos partes, la primera es una construcción de los cinturones de confianza combinando la construcción de Neymann y la regla de ordenación de Feldman-Cousins, la segunda es la construcción de curvas $1 - CL$ (CL : confidence level por sus siglas en inglés) usando la ecuación (4-11) como prueba estadística cuya distribución se compara con los datos medidos para obtener su p valor o $1 - CL$ valor, la diferencia entre estas dos partes es que la primera solo incluye mediciones singulares dentro de la función likelihood, esto es, el tratamiento del algoritmo es dato por dato, la segunda es posible incluir ensembles o vectores de datos además de poder incluir parámetros nuisance en la función likelihood y a partir de esto obtener la distribución de R por medio de la técnica Toys MC. En ambas construcciones se puede elegir con libertad el nivel de confianza $1 - \alpha$ deseado. Es necesario realizar la prueba de cobertura para corroborar que esta sea aproximada al nivel de confianza definido, logrando esto, es posible concluir que los intervalos de confianza generados tienen buena confiabilidad.

4.4.1. Cinturón de Confianza

Para construir los cinturones de confianza de Feldman-Cousins por medio del algoritmo de Karbach empezamos por tomar el negativo del logaritmo natural de la ecuación (4-11), esto con el objetivo de reducir la cantidad de cómputo, a este nuevo objeto lo definiremos como $\Delta\chi_{\text{toys}}^2$

$$-2 \ln R \equiv \Delta\chi_{toys}^2 = -2 \ln \mathcal{L}(\mu, x) - 2 \ln \mathcal{L}(\hat{\mu}, x) \quad (4-17)$$

$$= \chi^2(\mu, x) - \chi^2(\hat{\mu}, x) \quad (4-18)$$

Notar que x es una cantidad singular, no un vector de datos. El algoritmo para el caso de una Gaussiana es el siguiente:

- Para un valor parámetro verdadero μ_0 se generan n eventos x_{toy} que siguen la densidad de probabilidad $G(\sigma, \mu = \mu_0)$ con un σ conocido.
- Para cada evento generado x_{toy} , se obtiene $\Delta\chi_{toys}^2$ aplicando la restricción $\hat{\mu} = \max[0, x]$.
- Encontrar el valor $k_\alpha = \Delta\chi_c^2$ tal que una fracción de los eventos x_{toy} cumpla que $\Delta\chi_{toys}^2 < \Delta\chi_c^2$.
- Obtener el intervalo de aceptación $[x_{inf}, x_{sup}]$ éste es dado por todos los valores de x_{toy} tal que $\Delta\chi_{toys}^2 < \Delta\chi_c^2$.

La construcción queda completa cuando se itera el algoritmo variando solamente el parámetro verdadero μ_0 sobre un intervalo de interés $[\mu_0, \mu_\eta]$.

De la ecuación (4-17) vemos como los efectos de la frontera se vuelven visibles, esto es, para valores $x < 0$ la curva $\Delta\chi_{toys}^2$ tiende convertirse en una recta dado por $\Delta\chi_{toys}^2 = (x - \mu)^2 - x^2 = \mu^2 - 2\mu x$ y para la región $x > 0$ sigue una parábola $\Delta\chi_{toys}^2 = (x - \mu)^2$.

Validación del Algoritmo

Con el objetivo de validar este algoritmo se desarrollan los pasos anteriores y se comparan los resultados con los resultados originales de Feldman-Cousins, además se obtiene la probabilidad de cobertura. El algoritmo anterior se implementa con los siguientes valores:

- Consideremos una densidad de probabilidad Gaussiana con $G(x | \sigma = 1, \mu)$ con un σ conocido.
- $n = 10000$ eventos.
- Un intervalo dado por $[\mu_0 = 0.01, \dots, \mu_\eta = 8.01]$ con pasos $\Delta\mu = 0.01$.
- $1 - \alpha = 0.90$.

Los resultados gráficos del algoritmo se muestran en las Fig. 4-7 y 4-8.

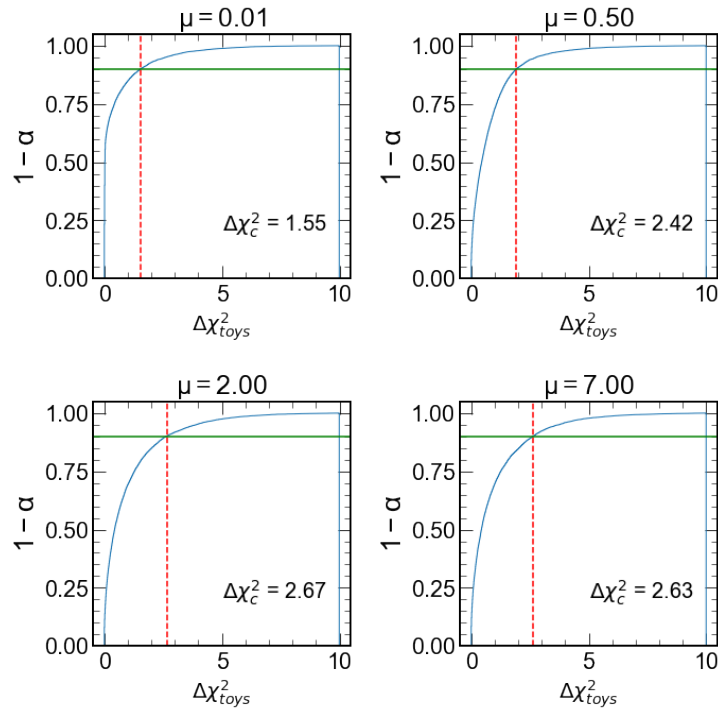


Figura 4-7.: Función cumulativa de $\Delta\chi^2_{toys}$ (curva azul) y $\Delta\chi^2_c$ (línea roja vertical). La cantidad $\Delta\chi^2_c$ queda determinada por el nivel de confianza elegido $1 - \alpha$ (línea horizontal verde), en este caso 0.90.

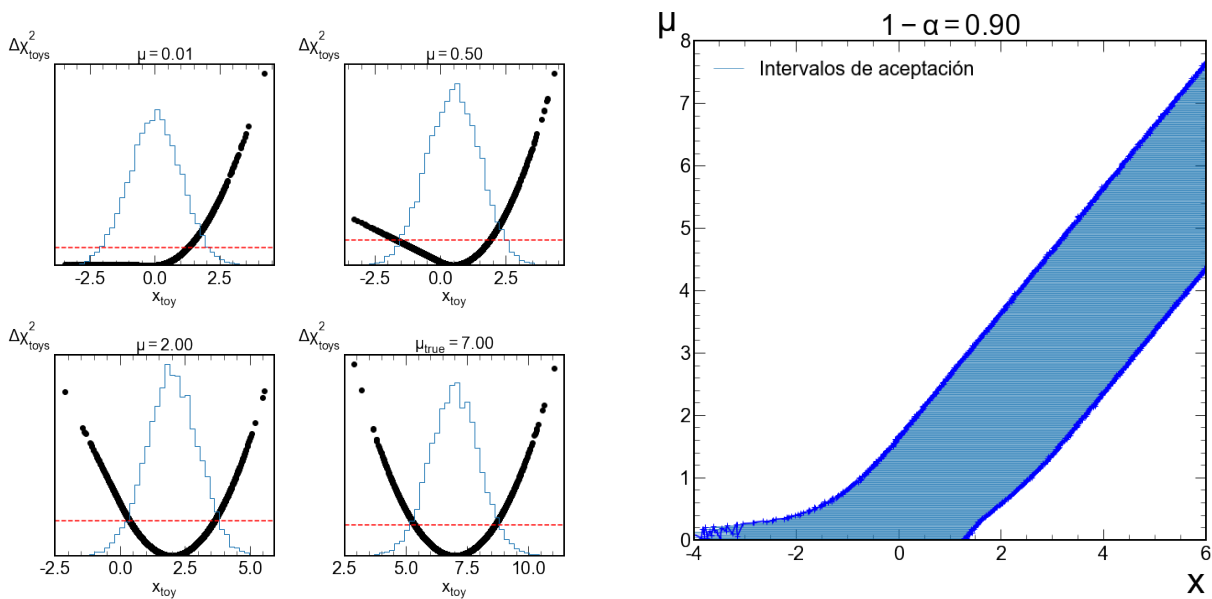


Figura 4-8.: (Izquierda) Distribución de x_{toy} (histograma), curva $\Delta\chi^2_{toys}$ (curva negra) y $\Delta\chi^2_c$ (recta horizontal roja). (Derecha) Cinturón de confianza resultante.

x_{obs}	Algoritmo de Karbach	Feldman-Cousins
-1.00	[0.00, 0.81]	[0.00, 0.81]
-0.50	[0.00, 1.17]	[0.00, 1.18]
0.00	[0.00, 1.65]	[0.00, 1.64]
0.50	[0.00, 2.15]	[0.00, 2.15]
1.00	[0.00, 2.64]	[0.00, 2.64]
1.50	[0.22, 3.15]	[0.22, 3.14]
2.00	[0.58, 3.64]	[0.58, 3.64]
2.50	[0.96, 4.12]	[0.95, 4.14]
3.00	[1.37, 4.66]	[1.37, 4.64]

Tabla 4-1.: Intervalos de confianza obtenidos al aplicar el algoritmo de Karbach para diferentes valores de x_{obs} a un nivel de confianza de 0.90.

De la Fig. 4-8 (izquierda) vemos como se transita de intervalos superiores a intervalos centrales dependiendo de la forma de la curva definida por $\Delta\chi_{toys}^2$, también se puede observar como se obtienen los intervalos de aceptación que son determinados por el conjunto de valores x_{toy} que están abajo de línea recta horizontal roja dado por $\Delta\chi_c^2$. Esta transición natural entre límites superiores e intervalos centrales es la razón por la cual la técnica de Feldman-Cousins también se le conoce como “enfoque unificado” De la figura derecha de 4-8 se observa la forma característica de la curva de Feldman-Cousins a diferencia de una construcción clásica de Neymann.

La tabla 4-1 muestra los intervalos de confianza que se obtuvieron del algoritmo de Karbach y junto a los reportados por Feldman Cousins. De la tabla 4-1 podemos observar como al aumentar los valores medidos u observados $x_0 = x_{obs}$ pasamos de intervalos de confianza con límites superiores $[0, \mu_{sup}]$ a intervalos centrales $[\mu_{inf}, \mu_{sup}]$, los resultados coinciden con los valores reportados por Feldman-Cousins.

Lo último de esta validación del algoritmo es obtener la cobertura para todos los valores del parámetro μ . El procedimiento fue el siguiente, para cada uno de los valores del intervalo $[\mu_0 = 0.01, \dots, \mu_\eta = 8.01]$ se generaron $I_T = 10000$ intervalos de confianza y se contaron el número total de estos intervalos que contuvieron al parámetro μ y esto se repite para cada μ del intervalo, con esta información se aplica el cociente dado por la probabilidad de cobertura (4-3). Los resultados se muestran en la Tabla 4-2 y en la Figura 4-9.

μ	$I_T = 10^4$	$I_T = 10^5$	$I_T = 10^6$	$I_T = 10^7$
0.1	0.9036	0.8985	0.8994	0.8995
0.5	0.9038	0.8975	0.8993	0.8991
1.0	0.9105	0.9064	0.9070	0.9066
3.0	0.9015	0.9010	0.9021	0.9020
5.0	0.8938	0.8985	0.8981	0.8984
8.0	0.8946	0.9014	0.9026	0.9025

Tabla 4-2.: Comportamiento asintótico de $\beta \rightarrow 1 - \alpha$ (cuadrícula gris) para diferentes valores de $\mu \in I_T$.

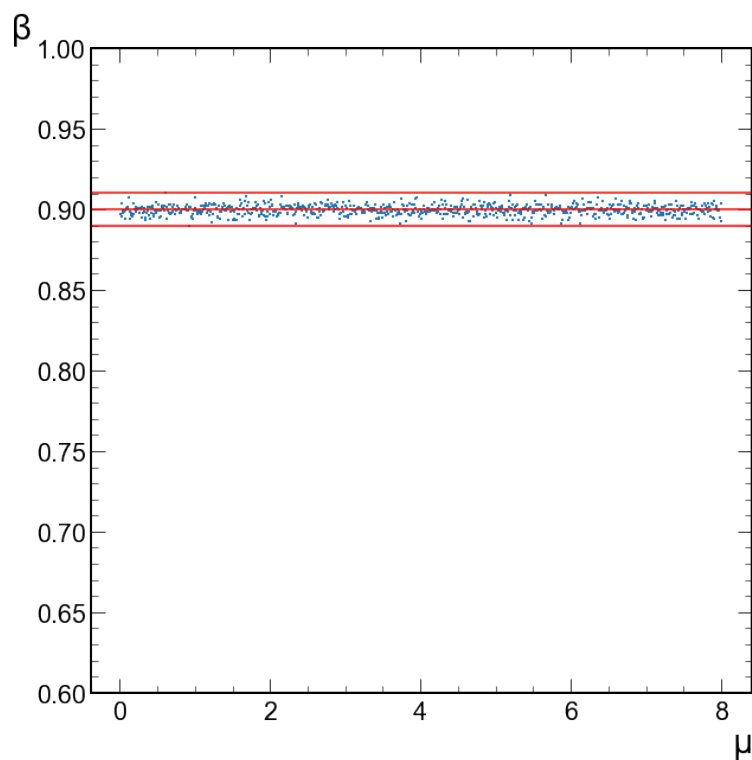


Figura 4-9.: Cobertura obtenida para $I_T = 10000$. Los resultados muestran fluctuaciones alrededor de $1 - \alpha = 0.90$.

De los resultados de la tabla 4-2 y de la probabilidad de cobertura $\beta \approx 1 - \alpha$ se concluye que el algoritmo de Karbach-Feldman-Cousins presentan coberturas con buena aproximación al nivel de confianza elegido, por tanto el algoritmo es válido.

4.4.2. Curvas $1 - CL$

El algoritmo Karbach-Feldman-Cousins para la generación de los cinturones de confianza nos sirve cuando consideramos un sólo parámetro de interés a un nivel de confianza específico, con esto es posible obtener los intervalos de confianza para cualquier medición que esté dentro del dominio de x , por otro lado, las curvas $1 - CL$ proporcionan intervalos de confianza para todos los niveles de confianza $1 - \alpha \in [0, 1]$ en una gráfica asociada a una medición x_{obs} , por lo que cada medición tiene su respectiva curva $1 - CL$. Estas curvas se construyen mediante una modificación al algoritmo anterior de Karbach-Feldman-Cousins:

- Al valor considerado del parámetro verdadero μ_0 , se genera una muestra aleatoria de n eventos x_{toy} de la densidad de probabilidad.
- Para cada evento x_{toy} se obtiene $\Delta\chi_{toys}^2$ con la restricción $\hat{\mu} = \max[0, x_{toy}]$.
- Se calcula $\Delta\chi_{datos}^2 \equiv \Delta\chi^2(x_{obs}, \mu_0)$ para el valor observado x_{obs} .
- El valor $1 - CL$ es dado por la fracción de eventos x_{toy} que tienen un mayor $\Delta\chi_{toys}^2$ que el valor de datos medidos $\Delta\chi_{datos}^2$:

$$1 - CL = \frac{N(\Delta\chi_{datos}^2 < \Delta\chi_{toy}^2)}{N_{toy}} \quad (4-19)$$

donde N_{toy} es el número total de eventos generados. El algoritmo anterior calcula solo un punto de la curva $1 - CL$, para obtener la curva completa es necesario iterar el algoritmo para todos los valores del parámetro μ de un intervalo de interés.

De nuevo desarrollando el procedimiento anterior para una Gaussiana obtenemos los resultados que se muestra en la Figura 4-10. Los intervalos de confianza asociados a las curvas $1 - CL$ para diferentes valores de x_{obs} se muestran en la Tabla 4-3

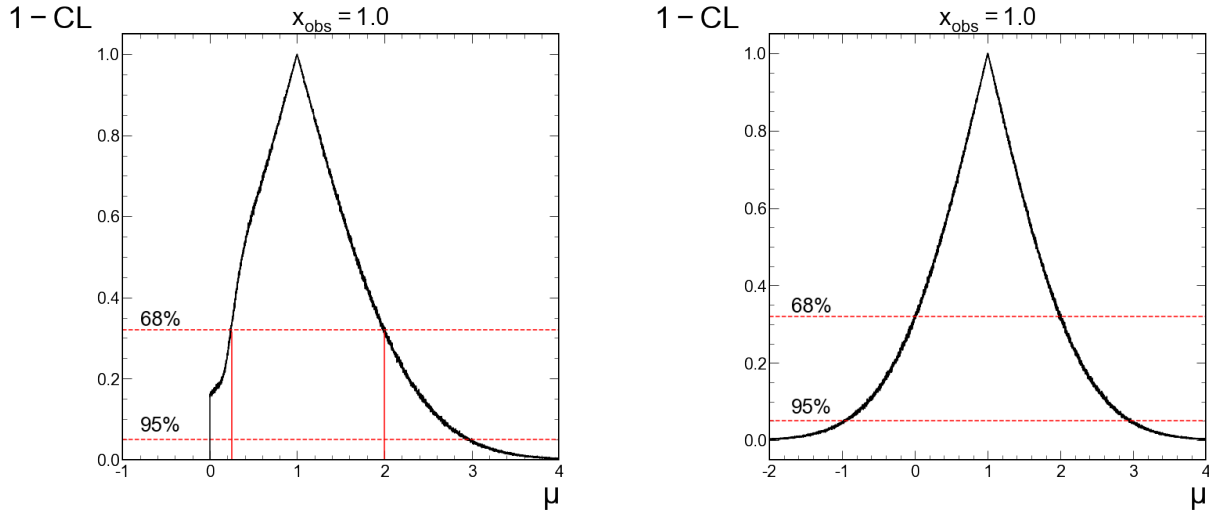


Figura 4-10.: (Izquierda) Curva $1 - CL$ para una Gaussiana con restricción $\hat{\mu} = \max[0, x_{toy}]$ para la observación $x_{obs} = 1.0$. (Derecha) Curva $1 - CL$ sin restricción. Los intervalos de confianza está dado por la intersección de las rectas horizontales con la curva negra.

	Algoritmo de Karbach		Feldman Cousins	
x_{obs}	68.27%	95%	68.27%	95%
0.00	[0.00, 1.00]	[0.00, 1.94]	[0.00, 1.00]	[0.00, 1.96]
0.10	[0.00, 1.10]	[0.00, 2.06]	[0.00, 1.10]	[0.00, 2.06]
0.50	[0.01, 1.50]	[0.00, 2.45]	[0.02, 1.50]	[0.00, 2.46]
1.00	[0.24, 1.99]	[0.00, 2.95]	[0.24, 2.00]	[0.00, 2.96]
1.50	[0.56, 2.51]	[0.00, 3.47]	[0.56, 2.50]	[0.00, 3.46]
2.00	[1.00, 3.01]	[0.34, 3.93]	[1.00, 3.00]	[0.35, 3.96]

Tabla 4-3.: Intervalos de confianza usando el método de curvas $1 - CL$ de Karbach. Los resultados coinciden con los reportados por Feldman-Cousins.

4.4.3. Curvas $1 - CL$ Algoritmo Extendido

El algoritmo definido para la obtención de las curvas $1 - CL$ de la sección anterior nos sirve si el tratamiento es dato por dato, en cambio, si tenemos un conjunto de datos con n eventos $\vec{x} = \{x_0, \dots, x_n\}$ es posible extender el algoritmo anterior, el cambio radica en la ecuación (4-17) en lugar de recibir un dato singular x recibe un vector de datos \vec{x} y R es considerado una prueba estadística[29], esto es, es posible obtener distribuciones de R de N muestras toy Monte Carlo, con esto, los datos medidos $\Delta\chi^2_{datos}$ se comparan con la distribución $\Delta\chi^2_{toys}$ y se obtiene el p o $1 - CL$ valor para cada parámetro que se itera. La ventaja de este enfoque es que podemos incluir

parámetros nuisance definidos en la ecuación (2-11), es decir, mediante el razón de likelihood de perfil

$$R = \frac{\mathcal{L}(\mu, \hat{\theta} | \vec{x})}{\mathcal{L}(\hat{\mu}, \hat{\theta} | \vec{x})} \quad (4-20)$$

$$\Delta\chi_{toys}^2 = -2 \ln \mathcal{L}(\mu, \hat{\theta} | \vec{x}) + 2 \ln \mathcal{L}(\hat{\mu}, \hat{\theta} | \vec{x}) \quad (4-21)$$

donde $\hat{\theta}$ es el valor del parámetro nuisance θ que maximiza la función likelihood para un μ fijo, $\hat{\theta}$ y $\hat{\mu}$ son los estimadores de máximo likelihood.

Los pasos resumidos del algoritmo extendido son los siguientes:

- A un valor de parámetro verdadero μ_0 generamos N muestras toys MC con n eventos de una densidad de probabilidad arbitraria.
- Calculamos $\Delta\chi_{toy}^2$ para cada uno de los N toy MC.
- Calculamos $\Delta\chi_{datos}^2(\vec{x}_{obs}, \mu_0)$ de los datos observados \vec{x}_{obs} .
- El valor $1 - CL$ es dado por la fracción de toys MC $\Delta\chi_{toys}^2$ mayor que el valor medido

$$1 - CL = \frac{N[\Delta\chi_{datos}^2(\mu_0, \vec{x}_{obs}) < \Delta\chi_{toy}^2(\mu_0, \vec{x})]}{N_{MC}} \quad (4-22)$$

donde N_{MC} es el número total de muestras toys MC generadas. La restricción del parámetro μ debe ser incluida en la minimización. Al igual que los algoritmos anteriores es necesario iterar para todos los valores del parámetro μ . El algoritmo anterior se puede visualizar por medio del diagrama de la Figura 4-11. Cabe destacar que los datos observados pueden ser datos reales, sin embargo aquí se usaron pseudo datos, además es necesario que al implementar el algoritmo se consideren las restricciones físicas cuando se minimiza $\Delta\chi_{toys}^2$ y $\Delta\chi_{datos}^2$.

Aplicando el algoritmo extendido a una Gaussiana con σ como parámetro nuisance y con las siguientes entradas:

- Intervalo de μ dado por $[1.00, 2.00]$ para pasos de $\Delta\mu = 0.005$.
- $N = 250$ toy MC con $n = 1000$ eventos por muestra.
- Nivel de confianza $1 - \alpha = 0.90$.
- Datos observados \vec{x}_{obs} obtenidos dado los parámetros $\mu_v = 1.5$ y $\sigma_v = 0.5$.

Las distribuciones $\Delta\chi_{toys}^2$ y la fracción $1 - CL$ para un valor de μ se muestran en la Figura 4-12.

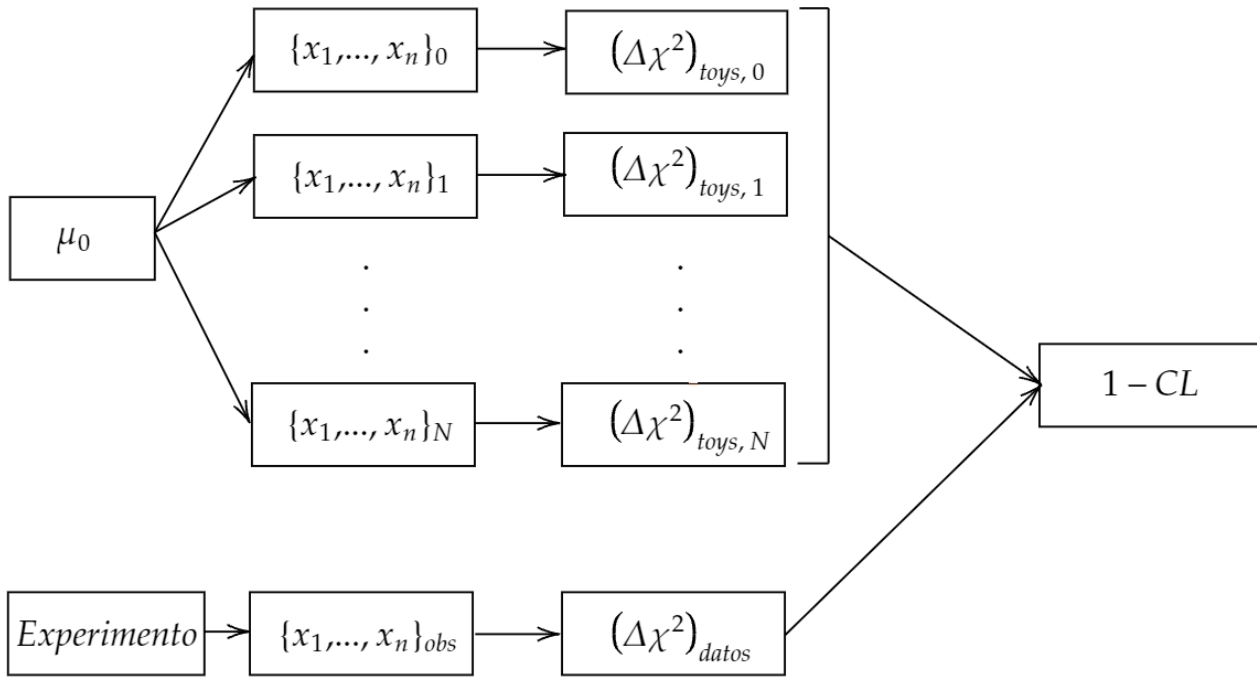


Figura 4-11.: Diagrama de operaciones del algoritmo extendido para un valor de $\mu = \mu_0$.

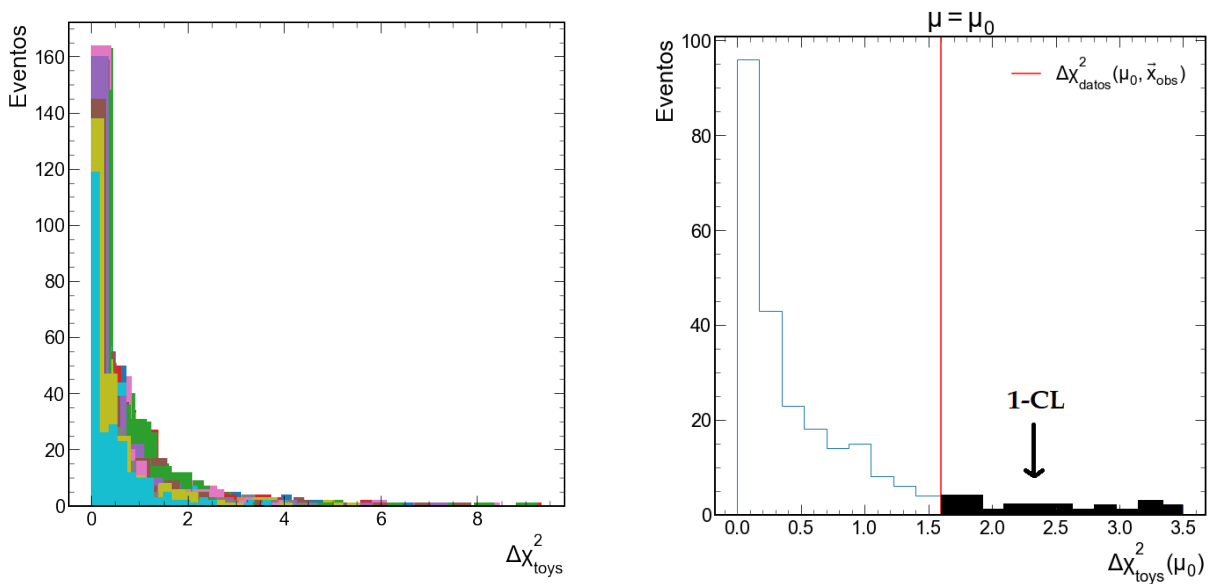


Figura 4-12.: (Izquierda) Distribuciones $\Delta\chi^2_{toys}$ obtenidas con el algoritmo extendido para todos los valores de μ del intervalo $[1.00, 2.00]$. (Derecha) Distribución de $\Delta\chi^2_{toys}(\mu_0)$ para un valor de $\mu = \mu_0$ y $\Delta\chi^2_{datos}$ con el cual se obtiene la fracción $1 - CL$ tal que $\Delta\chi^2_{datos} < \Delta\chi^2_{toys}$.

Iterando el algoritmo extendido para todos los valores de μ del intervalo $[1.00, 2.00]$, por cada iteración se obtiene un punto de la curva $1 - CL$, los resultados de todos los puntos conforman la curva $1 - CL$ como se muestra en la Figura 4-13 para dos conjuntos de datos observados.

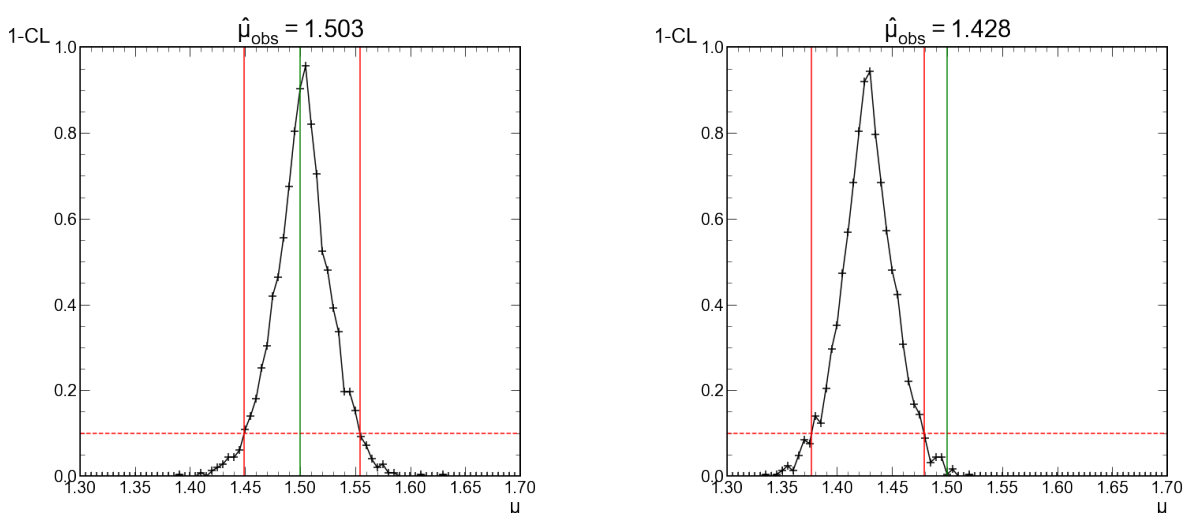


Figura 4-13.: (Izquierda) Curva $1 - CL$ para un conjunto de datos observados con estimador $\hat{\mu}_{obs} = 1.503$ cuyo intervalo de confianza contiene al parámetro verdadero $\mu_v = 1.5$. (Derecha) Curva $1 - CL$ con intervalo de confianza de datos observados que no cuyo intervalo de confianza no contiene al parámetro verdadero.

El máximo de las curvas $1 - CL$ se alcanza cuando el parámetro μ es aproximadamente igual al estimador $\hat{\mu}_{obs}$ de los datos observados \vec{x}_{obs} . La cobertura de este algoritmo se obtiene generando un conjunto de curvas $1 - CL$ para un parámetro verdadero. La cobertura β para 100 intervalos de confianza generados se muestran en la Tabla 4-2 para distintos valores del parámetro verdadero μ_v .

μ_v	β
1.2	0.88
1.4	0.90
1.5	0.88
1.6	0.93
1.8	0.89

Tabla 4-4.: Cobertura obtenida usando el algoritmo extendido para distintos valores de μ .

Los 100 intervalos de confianza (100 curvas $1 - CL$) se visualizan en la Figura 4-14, como se vio al inicio de esta sección la cobertura de los intervalos de confianza tiene que ser aproximadamente

igual al nivel de confianza con el cual se generaron dichos intervalos. De la Tabla 4-4 se concluye que los intervalos de confianza generados tienen una confiabilidad aproximadamente correcta.

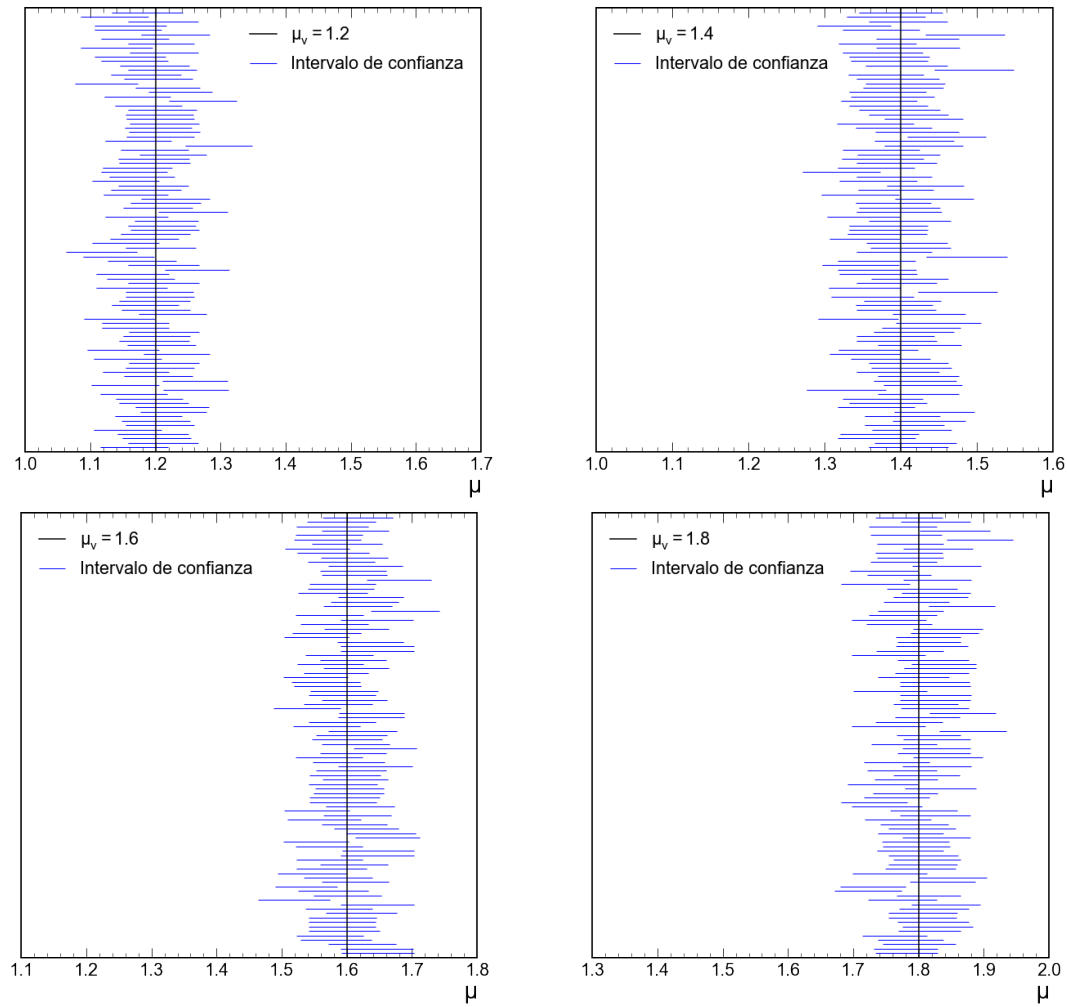


Figura 4-14.: Visualización de los intervalos de confianza generados con el algoritmo extendido que cubren o no al parámetro verdadero.

5. Aplicación y Resultados: Decaimiento

$$B^0 \rightarrow K_s^0 \mu^+ \mu^-$$

5.1. Estimación del Parámetro F_H

En el capítulo 1 se definió la distribución angular teórica asociada al decaimiento $B^0 \rightarrow K_s^0 \mu^+ \mu^-$, la densidad de probabilidad que describe la física total del sistema está dado por:

$$\begin{aligned} C(m, \cos(\theta) | F_H, Y_S, Y_B) &= E(Y_S)C_S(m, \cos(\theta) | F_H) + E(Y_B)C_B(m, \cos(\theta)) \\ &= E(Y_S)[S(m) \times D(\cos(\theta) | F_H)] + E(Y_B)[B_{ang}(m) \times B(\cos(\theta))] \\ &= E(Y_S)[CB(m) + G(m)] \times [A(\cos(\theta) | F_H) \times Ef(\cos(\theta))] + \\ &\quad E(Y_B)[E(m) + G(m)] \times [LB(\cos(\theta) + RB(\cos(\theta))] \quad (5-1) \end{aligned}$$

La construcción de la expresión (5-1) no es de interés para los resultados de esta tesis, las densidades de probabilidad que se usaron son las siguientes:

Densidad de probabilidad de señal $C_S(m, \cos(\theta) | F_H)$

$$S(m) = CB(m) + G(m) : \text{Masa de señal.}$$

$$D(\cos(\theta) | F_H) = A(\cos(\theta) | F_H) \times Ef(\cos(\theta)) : \text{Eficiencia del decaimiento.}$$

$$E(Y_s) : \text{Poisson (señal).}$$

$$CB(m) : \text{Bola de cristal.}$$

$$G(m) : \text{Gaussiana.}$$

$$A(\cos(\theta) | F_H) : \text{Distribución angular (ecuación 1-21).}$$

$$Ef(\cos(\theta)) : \text{Eficiencia.}$$

Densidad de probabilidad de fondo $C_B(m, \cos(\theta))$

$$\begin{aligned}
B(m) &= E(m) + G(m) : \text{Masa de fondo.} \\
B_{ang}(\cos(\theta)) &= LB(\cos(\theta)) + RB(\cos(\theta)) : \text{Fondo angular.} \\
E(Y_b) &: \text{Poisson (fondo).} \\
E(m) &: \text{Exponencial.} \\
G(m) &: \text{Gaussiana.} \\
LB(\cos(\theta)) &: \text{Polinomios de Bernstein (banda izquierda).} \\
RB(\cos(\theta)) &: \text{Polinomios de Bernstein (banda derecha).}
\end{aligned}$$

Este modelo describe la física del decaimiento $B^0 \rightarrow K_s^0 \mu^+ \mu^-$, a esta distribución la llamaremos densidad de probabilidad completa o total. El primer objetivo es aplicar los conceptos teóricos de los capítulos 2 y 3 con el fin de presentar los resultados de las estimaciones del parámetro F_H que sean consistentes y con sesgo despreciable.

Teniendo definida la densidad de probabilidad completa se generan $N = 2000$ muestras toy MC, de estas muestras obtenemos los N estimadores de F_H mediante el método de máximo likelihood usando la librería zfit para el ajuste y al optimizador de SLSQP en aplicación de la restricción $0 \leq F_H \leq 3$; la distribución de estimadores se grafica en un histograma y se procede a realizar un ajuste Gaussiano para obtener la media y la desviación estándar del estimador \hat{F}_H . Los resultados para un valor verdadero de $F_H = 0.2$ junto con su ajuste se muestran en la Figura 5-1.

La distribución de pulls de la Figura 5-1 da buenos resultados, ya que la media obtenida $\mu_p \approx 0$ y $\sigma_p \approx 1$, por tanto se concluye que la estimación del parámetro F_H es consistente y sin sesgo, el resultado que se reporta es

$$R_{pulls} = \frac{\mu_p - 0}{\epsilon_p} = \frac{0.0001 - 0}{0.0022} = 0.045 \quad (5-2)$$

donde ϵ_p es el error del ajuste Gaussiano de pulls, del resultado (5-2) se concluye que el valor está a 0.045σ del valor que debe salir. De la distribución de estimadores el valor que se reporta es

$$R_d = \frac{E[\hat{F}_H] - F_H}{\epsilon_d} = \frac{0.2025 - 0.2}{0.0022} = 1.1364 \quad (5-3)$$

donde ϵ_d es el error de ajuste de la distribución de estimadores.

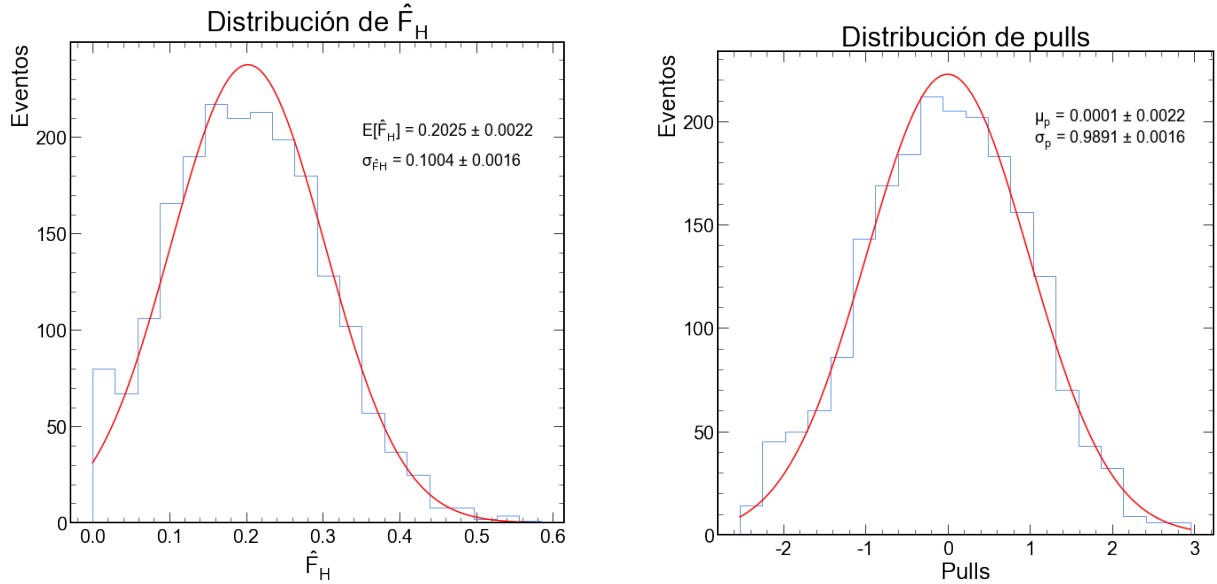


Figura 5-1.: Estimación del parámetro F_H para $N = 2000$ toy MC junto con sus errores asociados al ajuste Gaussiano.

5.2. Intervalos de Confianza y Cobertura de F_H

En esta sección se presentan los resultados que se obtuvieron al emplear el algoritmo extendido de Karbach-Feldman-Cousins del capítulo 4. La validación de la confiabilidad de los intervalos de confianza queda definido por la cobertura de las curvas $1 - CL$.

5.2.1. Curvas 1-CL para F_H

Para obtener la curvas $1 - CL$ asociadas al parámetro F_H de un conjunto de datos $\vec{x} = \{x_0, \dots, x_n\}$ partimos de la función likelihood asociada densidad de probabilidad completa (5-1)

$$\mathcal{L}(F_H, Y_s, Y_b | \vec{x}) = \prod_{i=1}^n C(m_i, \cos \theta_i | F_H, Y_s, Y_b) \quad (5-4)$$

consideramos a F_H como parámetro de interés y a los yields Y_s y Y_b como parámetros nuisance, con esto la razón de likelihood de perfil queda

$$R = \frac{\mathcal{L}(F_H, \hat{Y}_s(F_H), \hat{Y}_b(F_H) | \vec{x})}{\mathcal{L}(\hat{F}_H, \hat{Y}_s, \hat{Y}_b | \vec{x})} \quad (5-5)$$

Habiendo definido nuestra función R , se construyen las curvas $1 - CL$ para un intervalo de $F_H \in [0.00, 3.00]$ con pasos $\Delta F_H = 0.05$ y de 500 muestras Toys MC para cada valor F_H del intervalo

anterior, implementando paso a paso el algoritmo extendido de la sección 4.4.3 obtenemos las curvas $1 - CL$ para un parámetro verdadero $F_H = 0.2$ y un nivel de confianza de $1 - \alpha = 0.68$ que se muestran en la Figura 5-2 y Figura 5-3.

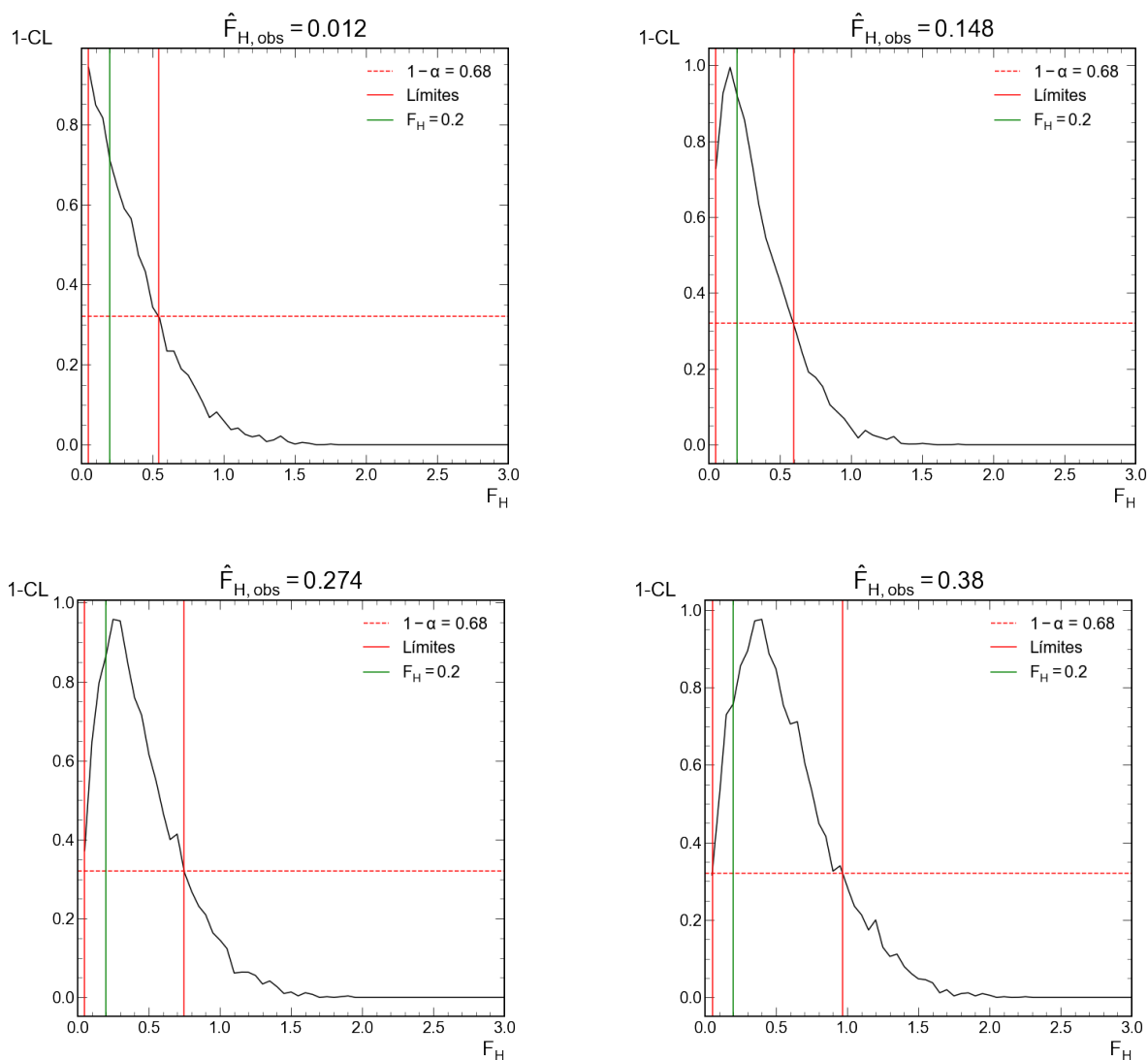


Figura 5-2.: Curvas $1 - CL$ e intervalos de confianza (límites) del parámetro F_H que contienen al parámetro verdadero $F_H = 0.2$.

Es necesario resaltar que los datos observados $\{x_0, \dots, x_n\}_{obs}$ que se usaron para obtener $\Delta\chi_{datos}^2$ fueron generados con pseudo datos, es decir, en esta tesis no se usaron datos reales.

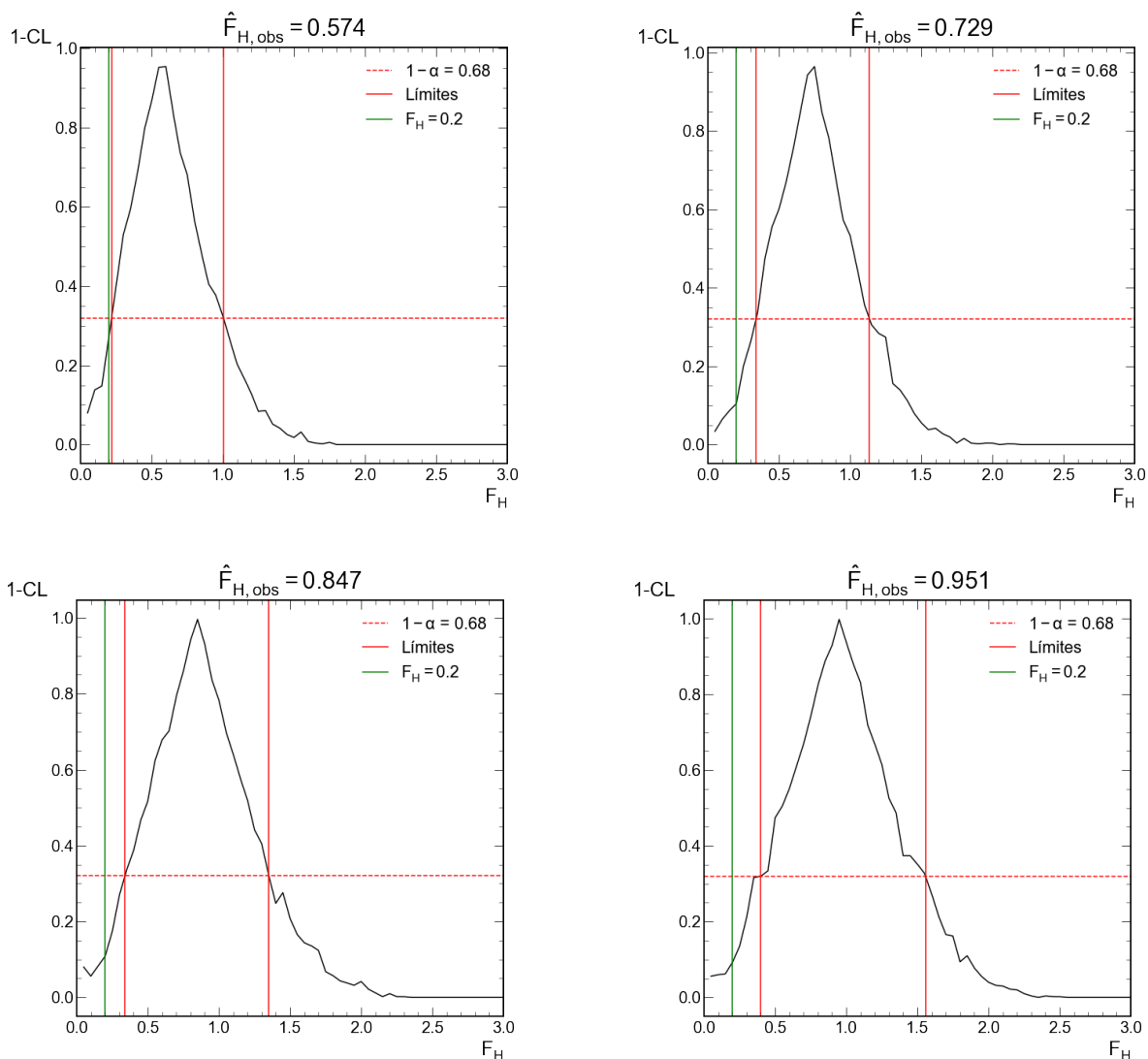


Figura 5-3.: Curvas $1-CL$ e intervalos de confianza (límites) del parámetro F_H que no contienen al parámetro verdadero $F_H = 0.2$.

Los intervalos de confianza numéricos de las gráficas asociadas a la Figura 5-2 y Figura 5-3 se presentan en la Tabla 5-1.

Ya que no todos los intervalos de confianza contienen al parámetro verdadero es necesario validar los intervalos de confianza generados, se procede a obtener la cobertura de los 500 intervalos de confianza obtenidos de cada una de las curvas $1-CL$, la totalidad de los intervalos de confianza para distintos valores de nivel de confianza se visualizan en la Figura 5-4 y Figura 5-5.

Intervalos de confianza para $F_H = 0.2$				
$\hat{F}_{H,obs}$	68%	90%	95%	99%
0.012	[0.000, 0.543]	[0.000, 0.860]	[0.000, 1.023]	[0.000, 1.433]
0.148	[0.000, 0.595]	[0.000, 0.867]	[0.000, 0.988]	[0.000, 1.333]
0.274	[0.000, 0.749]	[0.000, 1.069]	[0.000, 1.264]	[0.000, 1.625]
0.380	[0.051, 0.969]	[0.000, 1.369]	[0.000, 1.493]	[0.000, 1.862]
0.574	[0.219, 1.005]	[0.067, 1.233]	[0.000, 1.360]	[0.000, 1.596]
0.729	[0.340, 1.137]	[0.189, 1.421]	[0.077, 1.517]	[0.000, 1.825]
0.847	[0.340, 1.351]	[0.185, 1.721]	[0.000, 1.829]	[0.000, 2.110]
0.951	[0.400, 1.556]	[0.210, 1.866]	[0.000, 1.969]	[0.000, 2.250]

Tabla 5-1.: Intervalos de confianza numéricos para unos estimadores $\hat{F}_{H,obs}$ obtenidos a partir del parámetro verdadero $F_H = 0.2$.

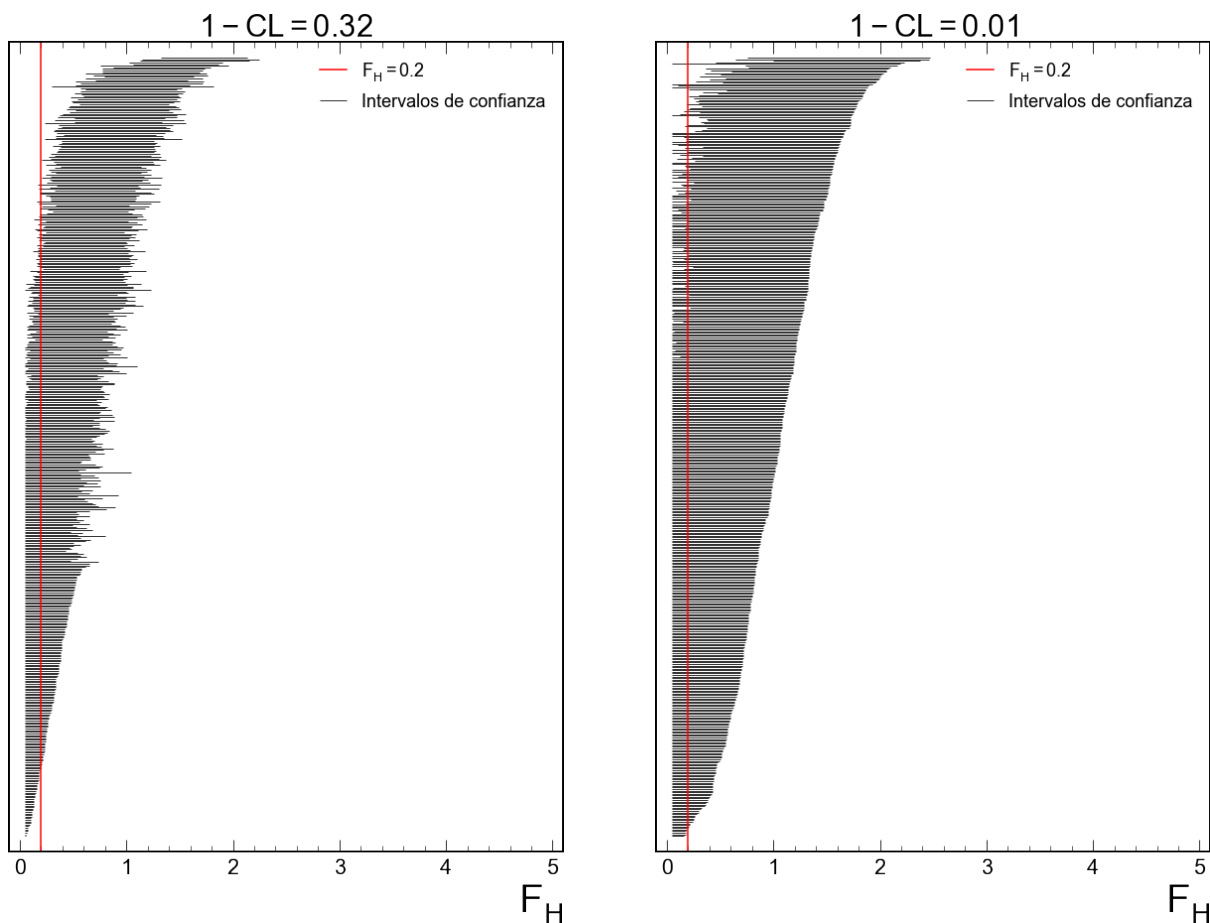


Figura 5-4.: Visualización de los intervalos de confianza obtenidos de cada una de las 500 curvas $1 - CL$ generadas para los niveles de confianza de 68% y 90%.

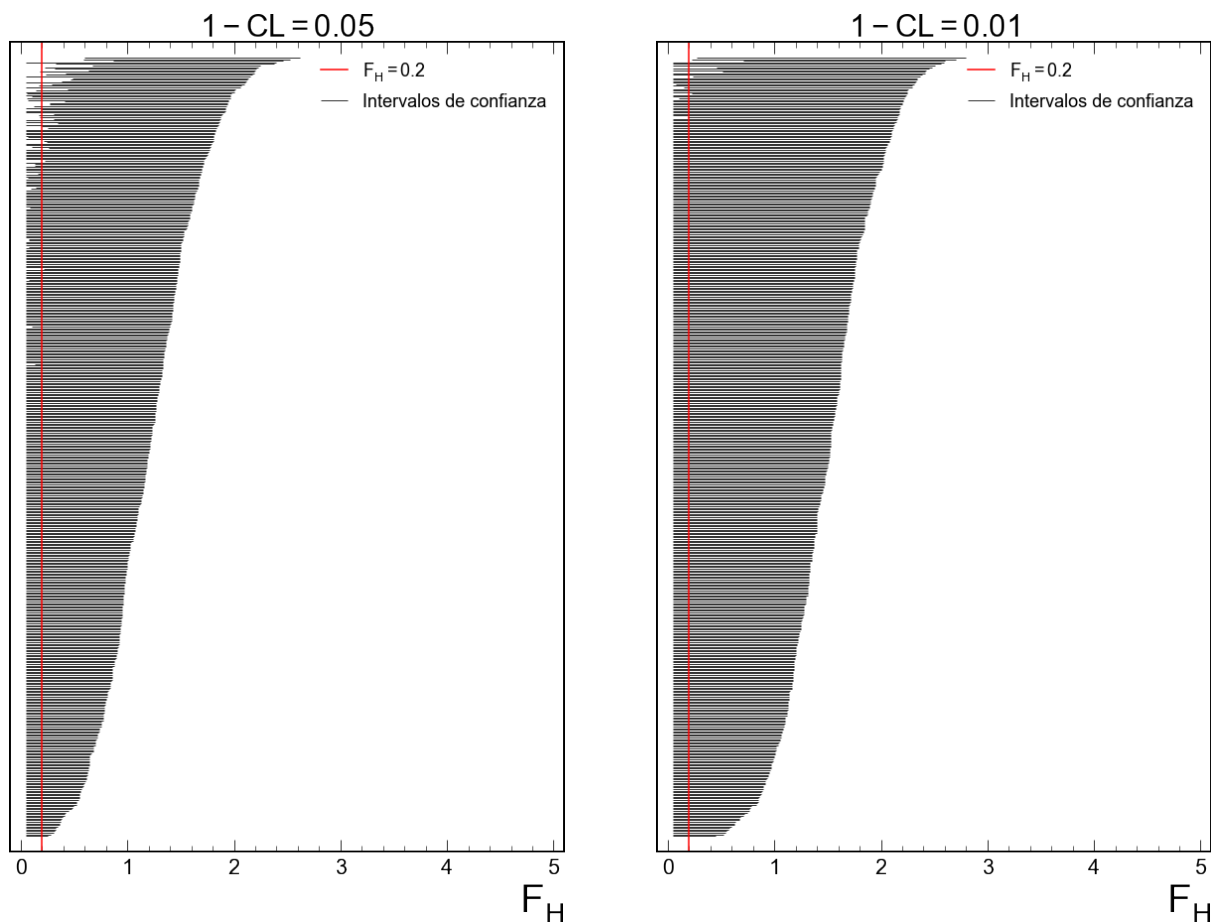


Figura 5-5.: Visualización de los intervalos de confianza obtenidos de cada una de las 500 curvas $1 - CL$ generadas para los niveles de confianza de 95% y 99%.

Ya que no todos los intervalos de confianza contienen al parámetro verdadero, se presenta la cobertura numérica, esta se obtiene de la razón de intervalos de confianza que contienen al parámetro verdadero $F_H = 0.2$ entre el número total de intervalos generados, estos resultados se muestran en la Tabla 5-2.

Cobertura $F_H = 0.2$	
$1 - \alpha$	β
0.68	0.69
0.90	0.89
0.95	0.94
0.99	0.99

Tabla 5-2.: Coberturas obtenidas de todos los intervalos de confianza generados.

5.3. Conclusiones

La estimación de parámetros e intervalos de confianza realizados en esta tesis por medio de la generación de pseudo datos aleatorios Toys MC para valores cercanos a la frontera del parámetro F_H nos permitieron presentar resultados con buenas propiedades estadísticas asociadas a la distribución angular de partículas del decaimiento $B^0 \rightarrow K_s^0 \mu^+ \mu^-$. La estimación paramétrica fue consistente y se obtuvo sin sesgo, lo que validó la técnica de ajuste utilizada. Se logró reproducir el cinturón de confianza de Feldman-Cousins con buena aproximación para el caso de una Gaussiana y la cobertura obtenida fue aproximadamente igual al nivel de confianza definido. Los intervalos de confianza para F_H obtenidos de las curvas $1 - CL$ por medio del algoritmo extendido y frecuentista de Karbach-Feldman-Cousins presentaron su característica de estimador, se logró corroborar que la cobertura fuera aproximadamente igual a los niveles de confianza definidos para un valor del parámetro F_H verdadero cercano a la frontera, por tanto se concluye que los intervalos de confianza resultantes del algoritmo extendido aplicado al parámetro F_H son válidos para una muestra representativa de 500 curvas $1 - CL$.

A. zfit y SLSQP

En física experimental de altas energías el lenguaje de programación que usualmente se usa es el C++ y el modelado estadístico es hecho con la herramienta ROOT/RooFit. En Python, la librería zfit [27] está hecha para el ajuste de modelos estadísticos, es decir, su objetivo básico es maximizar la concordancia de un conjunto de datos a una densidad de probabilidad descrita por un conjunto de parámetros. Los componentes básicos son los siguientes:

- Modelo estadístico.
- Datos.
- Función de costo (Likelihood).
- Minimizador.
- Resultados.

El uso básico es el siguiente:

- Definición de los límites del observable de interés o variable aleatoria del modelo.

```
obs = zfit.Space(obs = "x", limits = (0,10))
```

- Importación de datos (por ejemplo de numpy).

```
datos = zfit.Data.from_numpy(array = data_np, obs = obs)
```

- Creación del modelo estadístico (una gaussiana como ejemplo).

```
mu = zfit.Parameter("mu", value = 1.0)\  
sigma = zfit.Parameter("sigma", value = 0.5)\  
gauss = zfit.pdf.Gauss(obs=obs, mu=mu, sigma=sigma)
```

- Función de costo.

```
nll = zfit.loss.UnbinnedNLL(model = gauss, data = datos)
```

- Minimizador (Minuit).

```
minimizador = zfit.minimize.Minuit()
```

- Resultado.

```
resultado = minimizador.minimize(nll)
```

El objeto *resultado* contiene toda la información del ajuste, de aquí se obtienen los estimadores e incertidumbres asociadas al estimador. Este procedimiento se puede extender para modelos estadísticos más complejos, sin embargo las componentes siguen siendo las mismas.

Cuando se requiere hacer minimizaciones con restricciones impuestas por la física del sistema, la paquetería SLSQP se puede integrar a las componentes básicas de zfit con el objetivo de obtener resultados físicamente válidos, por ejemplo, si se requiere restringir los valores del parámetro μ (usado en el modelo Gaussiano anterior) al intervalo $[0 \leq \mu \leq 3]$, el minimizador de SLSQP se usa con la siguiente líneas de código:

```
constriccion = ({'type': 'ineq', 'fun': lambda x: x[0]},\ \
{'type': 'ineq', 'fun': lambda x: x[0]-3})\ \
SLSQP_FULL = SLSQP_zfit.SLSQP(constraints=constriccion)
```

Bibliografía

- [1] T. M. Karbach, “Feldman-cousins confidence levels-toy mc method,” *arXiv preprint arXiv:1109.0714*, 2011.
- [2] G. J. Feldman and R. D. Cousins, “Unified approach to the classical statistical analysis of small signals,” *Physical Review D*, vol. 57, no. 7, p. 3873, 1998.
- [3] P. B. Pal, *An introductory course of particle physics*. Taylor & Francis, 2014.
- [4] W. N. Cottingham and D. A. Greenwood, *An introduction to the standard model of particle physics*. Cambridge university press, 2007.
- [5] Y. Nir, “Flavour physics and cp violation,” *arXiv preprint arXiv:1010.2666*, 2010.
- [6] G. Isidori, “Flavour physics and implication for new phenomena,” *The Standard Theory of Particle Physics: Essays to Celebrate CERN’s 60th Anniversary*, pp. 339–355, 2016.
- [7] S. Tsiskaridze, *Search for Flavor-Changing Neutral Current Top Quark Decays $t \rightarrow Hq$, with $H \rightarrow b\bar{b}$ in pp Collisions at $\sqrt{s} = 8$ TeV with the ATLAS Detector*. Springer, 2017.
- [8] S. Dawson, “Introduction to electroweak symmetry breaking,” *arXiv preprint hep-ph/9901280*, 1999.
- [9] A. Ali, “Flavour-changing-neutral-current processes in b physics,” *Nuclear Physics B-Proceedings Supplements*, vol. 39, no. 2-3, pp. 408–425, 1995.
- [10] F. Archilli, M.-O. Bettler, P. Owen, and K. Petridis, “Flavour-changing neutral currents making and breaking the standard model,” *Nature*, vol. 546, no. 7657, pp. 221–226, 2017.
- [11] A. J. Buras, *Gauge Theory of Weak Decays: The Standard Model and the Expedition to New Physics Summits*. Cambridge University Press, 2020.
- [12] F. Lehner, *Flavor-changing Neutral Currents at Hadron Colliders*. PhD thesis, Verlag nicht ermittelbar, 2005.
- [13] F. Teubert, “Rare decays probing physics beyond the standard theory,” in *The Standard Theory of Particle Physics: Essays to Celebrate CERN’s 60th Anniversary*, pp. 357–370, World Scientific, 2016.

- [14] R. Aaij, B. Adeva, M. Adinolfi, A. Affolder, Z. Ajaltouni, J. Albrecht, F. Alessio, M. Alexander, S. Ali, G. Alkhazov, *et al.*, “Angular analysis of charged and neutral $B \rightarrow K\mu^+\mu^-$ decays,” 2014.
- [15] C. Bobeth, G. Hiller, and G. Piranishvili, “Angular distributions of $b \rightarrow kll$ decays,” *Journal of High Energy Physics*, vol. 2007, no. 12, p. 040, 2007.
- [16] L. Lyons, “Bayes and frequentism: a particle physicist’s perspective,” *Contemporary Physics*, vol. 54, no. 1, pp. 1–16, 2013.
- [17] G. Cowan, *Statistical data analysis*. Oxford university press, 1998.
- [18] J. Ocariz, “Probability and statistics for particle physicists,” *arXiv preprint arXiv:1405.3402*, 2014.
- [19] C. Maña, *Probability and Statistics for Particle Physics*. Springer, 2017.
- [20] S. Širca, *Probability for physicists*. Springer, 2016.
- [21] L. Lista, *Statistical methods for data analysis in particle physics*, vol. 941. Springer, 2017.
- [22] G. Ranucci, “The profile likelihood ratio and the look elsewhere effect in high energy physics,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 661, no. 1, pp. 77–85, 2012.
- [23] S. Algeri, J. Aalbers, K. D. Morå, and J. Conrad, “Searching for new phenomena with profile likelihood ratio tests,” *Nature Reviews Physics*, vol. 2, no. 5, pp. 245–252, 2020.
- [24] M. G. Kendall and S. Alan, “The advanced theory of statistics. vols. ii and iii.” 1961.
- [25] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics,” *The European Physical Journal C*, vol. 71, no. 2, pp. 1–19, 2011.
- [26] O. Behnke, K. Kröniger, G. Schott, and T. Schörner-Sadenius, *Data analysis in high energy physics: a practical guide to statistical methods*. John Wiley & Sons, 2013.
- [27] J. Eschle, A. P. Navarro, R. S. Coutinho, and N. Serra, “zfit: scalable pythonic fitting,” *SoftwareX*, vol. 11, p. 100508, 2020.
- [28] T. M. Karbach and M. Schlupp, “Constraints on yield parameters in extended maximum likelihood fits,” *arXiv preprint arXiv:1210.7141*, 2012.
- [29] J. Charles, S. Descotes-Genon, V. Niess, and L. V. Silva, “Modeling theoretical uncertainties in phenomenological analyses for particle physics,” *The European Physical Journal C*, vol. 77, no. 4, pp. 1–41, 2017.