

UNIVERSIDAD AUTÓNOMA METROPOLITANA
Unidad Iztapalapa

División de Ciencias Básicas e Ingeniería
Posgrado en Ciencias (Matemáticas)

"Pruebas de hipótesis secuenciales: aspectos computacionales de optimización y aplicaciones"

Tesis que presenta

M. en C. Fahil Farkhshatov

para obtener el grado de Doctor en Ciencias (Matemáticas)

Matrícula: 2212801602

ORCID: 0009-0005-5068-9963

Correo electrónico: failfarkshatov@gmail.com

Asesor:

Dr. Andrey Novikov

Jurado:

Presidente: Dr. José Andrés Christen Gracia

Secretario: Dr. Alejandro Román Vásquez

Vocal: Dr. Andrey Novikov

Vocal: Dr. Evgueni Gordienko

Vocal: Dr. Hugo Adán Cruz Suárez

Iztapalapa, Ciudad de México, 16 de Diciembre de 2025

Agradecimientos

En primer lugar, quiero expresar mi sincero agradecimiento a mi asesor, el Dr. Andrey Novikov, por su orientación, apoyo y generosidad al compartir sus conocimientos. Su experiencia y claridad en los temas de esta investigación fueron fundamentales para el desarrollo de este trabajo, y para mí ha sido un privilegio contar con su guía a lo largo de este proceso.

Quiero expresar también mi agradecimiento a México, país que me abrió sus puertas y me ofreció la oportunidad de formarme en un ambiente académico de gran calidad. Agradezco especialmente al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONACYT) por el apoyo financiero brindado durante mis estudios, el cual me permitió dedicarme plenamente a la investigación y al desarrollo de esta tesis.

Deseo expresar un agradecimiento muy especial a Andrej Novikov Jr., quien hizo posible mi primer acercamiento a este programa y me permitió conocer a mi asesor. Su apoyo, tanto en lo personal como en lo material, fue decisivo para que pudiera realizar este camino. Le agradezco profundamente su ayuda, sus consejos y la orientación que me brindó en los momentos más importantes de esta etapa. También agradezco a Petr Novikov por su apoyo moral a lo largo de este proceso.

Quiero expresar también mi agradecimiento a los amigos y personas que conocí en México, quienes me brindaron su apoyo y compañía durante mi estancia. Su ayuda fue esencial para mi adaptación y, en momentos decisivos, me ofrecieron incluso un lugar donde vivir, lo que me permitió continuar con mis estudios y concluir esta tesis. Estoy profundamente agradecido por su generosidad y amistad.

Finalmente, agradezco sinceramente a mi abuelita y a mi tía Nailya-apa por el apoyo que me brindaron en los inicios de este camino y por la confianza que depositaron en mí a lo largo de todo el proceso.

A todos, gracias.

Índice general

1. Pruebas secuenciales **2**

1.1. Introducción 2

1.2. Planes de muestreo óptimos 4

2. El problema de Kiefer-Weiss para familias exponenciales discretas **10**

2.1. Pruebas óptimas en el problema modificado de Kiefer-Weiss 10

2.2. Cotas para las regiones de continuación 12

2.3. Diseño computacional de pruebas óptimas 14

2.4. Característica de operación, número promedio muestral y fórmulas relacionadas 17

2.5. Resultados numéricos 20

2.5.1. Distribución Bernoulli 20

2.5.2. Familias exponenciales discretas 28

2.6. Análisis de los resultados numéricos 32

2.6.1. Distribución Bernoulli 32

2.6.2. Familias exponenciales discretas 35

3. El problema de Kiefer-Weiss para familias exponenciales continuas **37**

3.1. Solución del problema de Kiefer-Weiss modificado para familias exponenciales 37

3.2. Algoritmos para el cálculo de planes de muestreo óptimos 39

3.2.1. Cotas para las regiones de continuación 40

3.2.2. Algoritmos computacionales para el diseño de pruebas óptimas 41

3.2.3. Característica de operación, número promedio de observaciones y fórmulas relacionadas 43

3.3. Resultados numéricos 46

4. Pruebas secuenciales en grupos **59**

4.1. Planes de muestreo óptimos 59

4.1.1. Pruebas secuenciales en grupos 59

4.1.2. Pruebas secuenciales óptimas por grupos para familias exponenciales . . . 63
4.2. Algoritmos para el cálculo de planes de muestreo óptimos . . . 68
4.2.1. Cálculo de pruebas secuenciales óptimas por grupos . . . 68
4.2.2. Característica de operación, costo muestral esperado y fórmulas relacionadas . . . 72
4.3. Software complementario . . . 76
4.4. Aplicaciones y resultados numéricos . . . 78
4.4.1. Planes de muestreo óptimos para inspección por atributos . . . 79
4.4.2. Planes óptimos en ensayos clínicos de Fase II . . . 80
4.4.3. Planes secuenciales por grupos óptimos para la prueba de medias normales . . . 82

5. Conclusiones y perspectivas **94**

Bibliografía **97**

Índice de figuras

2.1. Desempeño de la prueba óptima para $\theta_0 = 0.05$ vs. $\theta_1 = 0.15$: $R(\psi^*)$ y $N(\theta_*; \psi^*)$ (arriba), $R_0(\psi^*)$ y $R_1(\psi^*)$ (abajo) 25
2.2. Desempeño de la prueba óptima para $\theta_0 = 0.2$ vs. $\theta_1 = 0.3$: $R(\psi^*)$ and $N(\theta_*; \psi^*)$ (arriba), $R_0(\psi^*)$ and $R_1(\psi^*)$ (abajo) 26
2.3. Desempeño de la prueba óptima para $\theta_0 = 0.45$ vs. $\theta_1 = 0.55$: $R(\psi^*)$ and $N(\theta_*; \psi^*)$ (arriba), $R_0(\psi^*)$ and $R_1(\psi^*)$ (abajo) 27

3.1. Prueba óptima de Kiefer-Weiss: número promedio de observaciones bajo θ 51
3.2. Prueba óptima de Kiefer-Weiss: FSS frente al número promedio de observaciones bajo θ 53
3.3. Prueba óptima de Kiefer-Weiss: FSS frente al número promedio de observaciones bajo H_0 54
3.4. Prueba óptima de Kiefer-Weiss: FSS frente al número promedio de observaciones bajo H_1 55
3.5. Prueba óptima de Kiefer-Weiss: FSS frente al percentil 99 del número de observaciones bajo θ 56
3.6. Prueba óptima de Kiefer-Weiss: θ como función de α y β 57
3.7. Prueba óptima de Kiefer-Weiss: FSS frente al número máximo de observaciones requerido 58

Índice de tablas

2.1. Distribución de Bernoulli con $\theta_0 = 0.05$, $\theta_1 = 0.15$ 23
2.2. Distribución de Bernoulli con $\theta_0 = 0.1$, $\theta_1 = 0.2$ 23
2.3. Distribución de Bernoulli con $\theta_0 = 0.2$, $\theta_1 = 0.3$ 24
2.4. Distribución de Bernoulli con $\theta_0 = 0.4$, $\theta_1 = 0.5$ 24
2.5. Distribución de Bernoulli con $\theta_0 = 0.45$, $\theta_1 = 0.55$ 25
2.6. Distribución de Poisson con media $\theta_0 = 0.5$ vs. $\theta_1 = 0.7$ 31
2.7. Distribución geométrica con media $\theta_0 = 1$ vs. $\theta_1 = 2$ 32
2.8. Distribución binomial con $\theta_0 = 0.05$ vs. $\theta_1 = 0.08$ y $n = 3$ 33
2.9. Comparación de pruebas para probabilidades de error asimétricas . . 34

3.1. Comparación entre las pruebas óptimas de Kiefer-Weiss, SPRT y FSST para distribución normal $\theta_0 = -0.1$, $\theta_1 = 0.1$ 52

4.1. Comparación con la norma ISO 2859-1 86
4.2. Parámetros de los planes optimizados en la Tabla 4.1 87
4.3. Planes optimizados para la Tabla 2 en Hu and Wang (2018b) 88
4.4. Parámetros de los planes optimizados en la Tabla 4.3 89
4.5. $\theta_0 = 0.05$ versus $\theta_1 = 0.2$, FSS= 37 90
4.6. $\theta_0 = 0.1$ versus $\theta_1 = 0.3$, FSS= 32 90
4.7. $\theta_0 = 0.2$ versus $\theta_1 = 0.4$, FSS= 44 91
4.8. $\theta_0 = 0.3$ versus $\theta_1 = 0.5$, FSS= 50 91
4.9. Eficiencia de los planes óptimos en el marco bayesiano. $\alpha = \beta = 0.05$ 92
4.10. Eficiencia de los planes óptimos para el problema de Kiefer-Weiss. $\alpha = \beta = 0.05$ 93

Resumen

En este trabajo se abordan problemas de pruebas de hipótesis en el marco del análisis estadístico secuencial, con especial atención a los problemas de Kiefer–Weiss. El objetivo principal es el diseño óptimo y la evaluación del desempeño de planes de muestreo para familias exponenciales, tanto del tipo continuo como discreto. Se desarrollaron algoritmos para la construcción de soluciones óptimas en el problema de Kiefer–Weiss, junto con fórmulas para calcular características de desempeño de las pruebas secuenciales, tales como la característica de operación, las probabilidades de errores, el número promedio de observaciones y otras medidas.

Se realizaron comparaciones numéricas de los planes óptimos obtenidos en el problema de Kiefer–Weiss con la prueba secuencial de razón de probabilidades (SPRT) y con la prueba de tamaño fijo, bajo los mismos niveles de error de los tipos I y II. Los algoritmos desarrollados se implementaron en el lenguaje de programación R, y el código correspondiente está disponible públicamente en un repositorio de GitHub (https://github.com/tosinabase/Kiefer-Weiss), con ejemplos para las distribuciones de Bernoulli, binomial, Poisson, binomial negativa, geométrica y normal.

En una segunda parte, se plantearon construcciones para el diseño de pruebas secuenciales por grupos bajo un criterio de costo esperado de muestreo, ponderado en un número finito de puntos del parámetro. Se desarrolló un método computacional que permite calcular los planes óptimos y sus características de desempeño. Los algoritmos correspondientes se implementaron en R, con código disponible en un repositorio público: https://github.com/tosinabase/groupsequential. Finalmente, se presentaron diversas aplicaciones tomadas de la literatura, en particular en el contexto de ensayos clínicos, que muestran mejoras sustanciales en eficiencia frente a planes conocidos previamente y confirman la utilidad práctica de los métodos propuestos.

Capítulo 1

Pruebas secuenciales

1.1. Introducción

Pruebas de hipótesis es un problema clásico de estadística matemática. Se trata de identificar una distribución a través de observaciones provenientes de ésta. El contexto clásico de las pruebas es una muestra de tamaño fijo, cuando, analizando los datos observados en la muestra se quiere decidir a favor de una de las dos hipótesis (contrastarlas, otro término que se usa para el procedimieto de la prueba es *contraste de hipótesis*). La calidad de la inferencia se puede medir a través de las probabilidades de error: la tipo I es el la probabilidad de rechazar la hipótesis base (nula) cuando ésta es verdadera, y la tipo dos de la probabilidad de aceptar la nula cuando se cumple la hipótesis alternativa. Un problema clásico (Neyman-Pearson) es minimizar la probabilidad de error tipo II manteniendo la probabilidad de error tipo I por debajo de un nivel dado (α). Determinación del tamaño de la muestra que permite hacerlo de tal manera que la probabilidad de error tipo II no exceda un nivel dado (β) siempre ha sido un problema importante de la estadística.

A mediados de los años cuarenta del siglo pasado, A. Wald propuso una prueba *secuencial* Wald (1945), en la cual los datos vienen apareciendo de uno a uno, los que permitió reducir sigificativamente el número *promedio* de observaciones que se necesitan para contrastar las hipótesis manteniendo los mismos niveles α y β de las probabilidades de error. Esta prueba se conoce hoy día como la prueba secuencial de razón de probabilidades (SPRT, por sus siglas en Inglés, Sequential Probability Ratio Tests). El resultado clásico de Wald and Wolfowitz (1948) dice que el número *promedio* de observaciones que usa la SPRT es mínimo, entre todas las pruebas secuenciales de niveles α y β de las probabilidades de error, y es mínimo tanto bajo la hipótesis nula como bajo la alternativa.

En este proyecto, pretendemos ver varios problemas relacionados con pruebas secuenciales de estadística, más específicamente, problemas *paramétricos*, cuando

las distribuciones hipotetizadas se identifican a través de un parámetro θ. Sea la hipótesis nula $H_0 : \theta = \theta_0$ y la alternativa $H_1 : \theta = \theta_1$ (tradicionalmente, las hipótesis *simples*, es decir las que están plenamente identificadas por los dos valores del parámetro θ_0 y θ_1 forman base para el desarrollo de métodos más generales).

El problema de optimización de las pruebas secuenciales en razón de probabilidades, Wald and Wolfowitz (1948) lo resuelven utilizando el método Bayesiano. Este resultado es teóricamente excelente porque se minimizan ambos números promedio de observaciones para las dos hipótesis a la vez. Pero si hay un error en la formulación del problema y realmente la hipótesis está entre esos puntos que suponemos (θ_0 y θ_1), el número promedio de observaciones no resulta ser mínimo, es más, en muchas ocasiones excede el número de observaciones que requiriera una prueba no secuencial (Neyman-Pearson) para las mismas α y β. Así surge el problema de Kiefer-Weiss.

Kiefer y Weiss (Kiefer and Weiss, 1957) demostraron teoremas estructurales sobre pruebas óptimas para varias formulaciones, incluido el problema de minimización del tamaño de muestra esperado $E_{\theta^*}\tau_\psi$ para un punto $\theta^* \in (\theta_0, \theta_1)$, el problema que se conoce hoy como problema de Kiefer-Weiss modificado. Bechhofer (1960) mostró la atractividad de minimizar el máximo tamaño de muestra esperado $\sup E_\theta\tau_\psi$ sobre todos los θ posibles. Kiefer and Weiss (1957) plantearon un enfoque para resolverlo. Y Weiss (1962) mostró cómo este problema, así se llama problema de Kiefer-Weiss, se reduce a problema de Kiefer-Weiss modificado en casos simétricos que involucran las distribuciones normal y binomial. Después Lai (1973) investigó el caso de proceso Wiener.

Freeman y Weiss (Freeman and Weiss, 1964) aplicaron la misma técnica a un caso más general del modelo de Bernoulli con hipótesis no simétricas; propusieron un esquema para encontrar una solución numérica aproximada al problema de Kiefer-Weiss haciendo $E_{\theta^*}\tau_\psi$ lo suficientemente cerca de $\sup_\theta E_\theta\tau_\psi$.

Lorden (1980) caracterizó la estructura de las pruebas óptimas en el problema de Kiefer-Weiss modificado para el caso particular de las familias de distribuciones de Koopman-Darmois uniparamétricas y mostró que, para $\theta_* \in (\theta_0, \theta_1)$, los tiempos de paro de estas pruebas están casi seguramente acotados, y dio este límite.

En la literatura se pueden encontrar muchos otros resultados relacionados con soluciones aproximadas o asintóticas del problema de Kiefer-Weiss, y sobre todo del problema de Kiefer-Weiss modificado. El punto de vista moderno sobre el problema de Kiefer-Weiss y otros relacionados se puede encontrar en la monografía Tartakovsky et al. (2014).

En los capítulos 1-3 propusimos un enfoque computacional para muestreo de familias exponenciales que proporciona solución numérica para los problemas de Kiefer-Weiss, incluyendo también, como, caso particular, los problemas clásicos Bayesianos y desarrollamos un paquete computacional para implementación de los

procedimientos.

En el presente proyecto también pretendemos estudiar las pruebas secuenciales por grupos, un enfoque que obtuvo mucha atención en la literatura en las últimas décadas, sobre todo en relación con las aplicaciones en ensayos clínicos. El enfoque secuencial de grupos (Pocock, 1977) ofrece una alternativa a las pruebas "puramente" secuenciales, que requiere la acumulación de datos para ser examinados cada vez que se registre una nueva observación. Jennison (1987) utilizó cálculos numéricos y derivó pruebas secuenciales grupales unilaterales óptimas para la media de una distribución normal con varianza conocida. Eales and Jennison (1992) derivaron pruebas unilaterales óptimas para hipótesis simétricas con probabilidades de error de tipo I y tipo II iguales. Chang (1996) obtuvo diseños óptimos bilaterales con regiones de aceptación y rechazo de la hipótesis nula en las etapas tempranas. Barber and Jennison (2002) extendieron las pruebas secuenciales grupales simétricas óptimas a la clase más amplia de diseños de pruebas secuenciales grupales unilaterales asimétricas. An and Chang (2016) propusieron un método de simulación para calcular pruebas secuenciales grupales óptimas que minimizan el número promedio de observaciones y cumplen con los requisitos de potencia y nivel de significación. En Hu and Wang (2018b) se propone un algoritmo de búsqueda orientado al uso computacional, que proporciona pruebas secuenciales por grupos casi óptimas para proporciones binarias. Las comparaciones numéricas de su eficiencia con algunos estándares nacionales e internacionales para procedimientos secuenciales por grupos de inspección por atributos muestran una alta eficiencia del método propuesto.

La propuesta principal de este parte del proyecto (Capítulo 4) es extender el enfoque computacional de los capítulos 1-3 a los problemas de pruebas secuenciales por grupos, desarrollando los métodos numéricos respectivos y implementándolos en forma de algoritmos y programas computacionales para sus aplicaciones en la práctica (en el primer plano, en ensayos clínicos).

1.2. Planes de muestreo óptimos

Sea $X_1, X_2, \ldots, X_n, \ldots$ un proceso estocástico a tiempo discreto.

Las v.a. X_i tienen su función de "densidad" de probabilidad conjunta

$$f_\theta^n = f_\theta^n(x_1, \ldots, x_n)$$

(en el sentido de la derivada de Radon-Nikodym de su distribución) con respecto a la medida producto y donde la distribución depende de un parámetro θ.

Consideramos el problema de pruebas de hipótesis simples $H_0 : \theta = \theta_0$ contra la alternativa $H_1 : \theta = \theta_1$.

Se tiene una *prueba estadística secuencial aleatorizada* (ψ, ϕ) donde ψ es una regla de paro y ϕ una regla de decisión.

Suponemos que

$$\psi = (\psi_1, \psi_2, \dots, \psi_n, \dots) \quad \text{y} \quad \phi = (\phi_1, \phi_2, \dots, \phi_n, \dots)$$

donde las funciones

$$\psi_n = \psi_n(x_1, x_2, \dots, x_n) \quad \text{y} \quad \phi_n = \phi_n(x_1, x_2, \dots, x_n)$$

son medibles y toman sus valores en $[0, 1]$, $n = 1, 2, 3, \dots$.

El experimento comienza con la obtención del valor x_1 de la primera observación de X_1 (etapa $n = 1$). Para cualquier etapa $n = 1, 2, \dots$, la función $\psi_n(x_1, x_2, \dots, x_n)$ puede ser interpretada como la probabilidad condicional de *parar el experimento y proceder a la toma de decisión*, dado que el experimento llegó a la etapa n y que las observaciones de $(X_1, \dots, X_n)$ hasta esta etapa fueron $(x_1, \dots, x_n)$. Si no hay paro, los experimentos continúan a la siguiente etapa y se toma una observación adicional x_{n+1} de X_{n+1}. Luego se aplica la regla ψ_{n+1} a $(x_1, x_2, \dots, x_n, x_{n+1})$ de la misma manera que antes, etc., hasta que el experimento finalmente se detiene.

Se supone que cuando el experimento se detiene, se debe tomar la decisión de aceptar o rechazar H_0. La función $\phi_n(x_1, x_2, \dots, x_n)$ es interpretada como probabilidad condicional de *rechazar* la hipótesis nula H_0, dado que el experimento se detiene en la etapa n donde $(x_1, \dots, x_n)$ es el vector de datos observados.

Para cualquier regla de paro ψ se definen las siguientes v. a.

$$c_n^{\psi} = c_n^{\psi}(x_1, \dots, x_n) = (1 - \psi_1)(1 - \psi_2) \dots (1 - \psi_{n-1}),$$

$$s_n^{\psi} = s_n^{\psi}(x_1, \dots, x_n) = c_n^{\psi} \psi_n,$$

la regla ψ genera la variable aleatoria τ_ψ llamada tiempo de paro, tal que $\tau_\psi \in \{1, 2, \dots\} \cup \{\infty\}$ y cuya distribución para $k \in \mathbb{N}$ está dada por:

$$P_\theta(\tau_\psi = k) = E_\theta s_n^{\psi}.$$

Para caracterizar el número de observaciones hasta obtener una decisión final, se trabaja una característica llamada el número promedio muestral (el número promedio de observaciones).

$$N(\theta; \psi) = E_\theta \tau_\psi = \sum_{n=1}^{\infty} n E_\theta s_n^{\psi}.$$

(cuando $\sum_{n=1}^{\infty} E_\theta s_n^\psi (= P_\theta(\tau_\psi < \infty)) = 1$, – en caso contrario es infinito).

Las probablidades de los errores de tipo I y tipo II para una prueba secuencial (ψ, ϕ) se define, respectivamente, como

$$\alpha(\psi,\phi) = \sum_{n=1}^{\infty} E_{\theta_0} s_n^\psi \phi_n \quad \text{y} \quad \beta(\psi,\phi) = \sum_{n=1}^{\infty} E_{\theta_1} s_n^\psi (1-\phi_n),$$

Las probabilidades de los errores se expresan convenientemente con característica de operación

$$\mathrm{OC}_\theta(\psi,\phi) = \sum_{n=1}^{\infty} E_\theta s_n^\psi (1-\phi_n), \tag{1.1}$$

tal que $\alpha(\psi,\phi) = 1 - \mathrm{OC}_{\theta_0}(\psi,\phi)$ y $\beta(\psi,\phi) = \mathrm{OC}_{\theta_1}(\psi,\phi)$.

Sea $\mathscr{S}(\alpha,\beta)$ el conjunto de todos los pruebas tales que

$$\alpha(\psi,\phi) \le \alpha, \quad \beta(\psi,\phi) \le \beta, \tag{1.2}$$

donde $\alpha, \beta \in [0,1]$ son unos números fijos.

El problema de encontrar las pruebas minimizando $\sup_\theta N(\theta;\psi)$ sobre todas las pruebas $\mathscr{S}(\alpha,\beta)$ es conocido como problema de Kiefer-Weiss (Kiefer and Weiss (1957), Bechhofer (1960), Weiss (1962), Lorden (1980))

$$\sup_\theta N(\theta;\psi) \xrightarrow[\mathscr{S}(\alpha,\beta)]{} \text{mín}\,. \tag{1.3}$$

El problema de Kiefer-Weiss modificado es de minimizar $N(\theta;\psi)$ sobre las todas $(\psi,\phi) \in \mathscr{S}(\alpha,\beta)$, para un valor fijo de θ (Kiefer and Weiss (1957), Weiss (1962))

$$N(\theta;\psi) \xrightarrow[\mathscr{S}(\alpha,\beta)]{} \text{mín}, \qquad \text{para un } \theta \text{ fijo.} \tag{1.4}$$

El problema en formulación bayesiana se define de la siguiente manera (Wald and Wolfowitz (1948))

$$\gamma_0 N(\theta_0;\psi) + \gamma_1 N(\theta_1;\psi) \xrightarrow[\mathscr{S}(\alpha,\beta)]{} \text{mín},$$

donde γ_0, γ_1 son números dados tales que $\gamma_0, \gamma_1 \ge 0$ y $\gamma_0 + \gamma_1 = 1$.

Las soluciones del problema modificado de Kiefer–Weiss pueden obtenerse, en una situación bastante general, mediante una variante del método de los multiplicadores de Lagrange.

Sea

$$L_\theta(\psi,\phi) = N(\theta;\psi) + \lambda_0 \alpha(\psi,\phi) + \lambda_1 \beta(\psi,\phi), \tag{1.5}$$

donde θ es un valor fijo del parámetro, y λ_0, λ_1 son constantes no negativas llamadas

multiplicadores de Lagrange.

Entonces, las pruebas que minimizan $N(\theta;\psi)$ bajo las restricciones descritas en (1.2), es decir, tales que $(\psi,\phi)\in\mathscr{S}(\alpha,\beta)$, pueden obtenerse mediante la minimización no restringida del funcional $L_\theta(\psi,\phi)$ sobre todas las pruebas (ψ,ϕ), mediante una elección adecuada de los multiplicadores de Lagrange ((Novikov, 2009b, Sección 2)).

En Lorden (1980) se muestra que, en el caso de observaciones i.i.d., el problema de minimizar la función de Lagrange (1.5) se reduce a un problema de paro óptimo para un proceso de Markov.

También es fácil ver que la búsqueda de pruebas bayesianas, como las utilizadas en Kiefer and Weiss (1957), es matemáticamente equivalente a la minimización del funcional $L_\theta(\psi,\phi)$.

Para construir pruebas óptimas es necesario hacer una suposición adicional sobre la estructura de la distribución de las observaciones. Sea $f_\theta^n = f_\theta^n(x_1,\dots,x_n)$ la derivada de Radon–Nikodym de la distribución conjunta de $(X_1,\dots,X_n)$ con respecto a una medida producto $\mu^n = \mu\otimes\cdots\otimes\mu$ (n veces), para $n=1,2,\dots$

En Lorden (1980) también se demuestra que, si las observaciones son i.i.d. y tienen una distribución perteneciente a una familia de Koopman–Darmois, entonces las pruebas que resuelven el problema modificado tienen tiempos de paro acotados con probabilidad uno, siempre que $\theta\in(\theta_0,\theta_1)$.

Describamos ahora la construcción de pruebas que minimizan la función de Lagrange evaluada en un cierto valor fijo de θ, dentro de la clase de pruebas *truncadas*, es decir, aquellas que no utilizan más de un número fijo H de observaciones (a este valor H también se le llama *horizonte*).

Formalmente, sea

$$\mathscr{S}^H = \{(\psi,\phi):\ c_{H+1}^\psi \equiv 0\}$$

la clase de todas esas pruebas.

La estructura de las pruebas que minimizan la función de Lagrange dentro de la clase $\mathscr{S}^H$ puede caracterizarse de la siguiente manera.

Definamos

$$V_{\theta,H}^H = \min\{\lambda_0 f_{\theta_0}^H,\ \lambda_1 f_{\theta_1}^H\}, \tag{1.6}$$

y recursivamente para $n = H-1,\dots,1$,

$$V_{\theta,n}^H = \min\left\{\lambda_0 f_{\theta_0}^n,\ \lambda_1 f_{\theta_1}^n,\ f_\theta^n + \mathcal{I}V_{\theta,n+1}^H\right\}, \tag{1.7}$$

donde

$$\mathcal{I}V_{\theta,n+1}^H(x_1,\dots,x_n) = \int V_{\theta,n+1}^H(x_1,\dots,x_{n+1})\,d\mu(x_{n+1}). \tag{1.8}$$

Obsérvese que las funciones $V^H_{\theta,n}$ definidas arriba, al igual que la función $L_\theta(\psi,\phi)$ en (1.5), dependen implícitamente de los multiplicadores λ_0 y λ_1.

nota 1 *Las funciones $V^H_{\theta,n}$ definidas anteriormente, al igual que la función $L(\psi,\phi)$ en (1.5), dependen implícitamente de los parámetros θ_0, θ_1, θ_*, así como de los multiplicadores λ_0 y λ_1.*

A partir de los resultados de la Sección 3.1 en Novikov (2009b), se obtiene fácilmente la siguiente caracterización de todas las pruebas secuenciales truncadas que minimizan la función $L(\psi,\phi)$.

Proposición 1.2.1 *Para toda prueba $(\psi,\phi)\in\mathscr{S}^H$ se cumple que*

$$L(\psi,\phi)\geq 1+\mathcal{I}V^H_{\theta,1}. \tag{1.9}$$

La igualdad en (1.9) se alcanza para una prueba $(\psi,\phi)\in\mathscr{S}^H$ si, y sólo si, satisface las siguientes dos condiciones:

a) *Para todo $n=1,2,\ldots,H-1$*

$$I_{\left\{\min\{\lambda_0 f^n_{\theta_0},\lambda_1 f^n_{\theta_1}\}<f^n_{\theta_*}+\mathcal{I}V^H_{\theta,n+1}\right\}}\leq\psi_n\leq I_{\left\{\min\{\lambda_0 f^n_{\theta_0},\lambda_1 f^n_{\theta_1}\}\leq f^n_{\theta_*}+\mathcal{I}V^H_{\theta,n+1}\right\}} \tag{1.10}$$

μ^n-c.t.p. sobre el conjunto $C^\psi_n=\{(x_1,\ldots,x_n): c^\psi_n(x_1,\ldots,x_n)>0\}$,

b) *Para todo $n=1,2,\ldots,H$*

$$I_{\left\{\lambda_0 f^n_{\theta_0}<\lambda_1 f^n_{\theta_1}\right\}}\leq\phi_n\leq I_{\left\{\lambda_0 f^n_{\theta_0}\leq\lambda_1 f^n_{\theta_1}\right\}} \tag{1.11}$$

μ^n-c.t.p. sobre el conjunto $S^\psi_n=\{(x_1,\ldots,x_n): s^\psi_n(x_1,\ldots,x_n)>0\}$.

Bajo condiciones muy generales, las pruebas óptimas no truncadas se obtienen a partir de los límites $V_{\theta,n}=\lim_{H\to\infty}V^H_{\theta,n}$. Las reglas de paro óptimas para las pruebas no truncadas se definen sustituyendo $V^H_{\theta,n}$ por $V_{\theta,n}$:

$$\psi_n=I_{\{\min\{\lambda_0 f^n_{\theta_0},\lambda_1 f^n_{\theta_1}\}\leq f^n_\theta+\mathcal{I}V_{\theta,n+1}\}}, \tag{1.12}$$

en la expresión (1.10) para todo n, mientras que las reglas de decisión se mantienen sin cambios, aplicando (1.11) para todo n (véase la Sección 3 en Novikov (2009b)).

Denotemos por $\mathscr{M}^H(\theta_*,\lambda_0,\lambda_1)$ la clase de todas las pruebas que satisfacen las condiciones a) y b) de la Proposición 1.2.1, y por $\mathscr{M}(\theta_*,\lambda_0,\lambda_1)$ la clase de todas las pruebas (no truncadas) que satisfacen la condición (1.10) con $V^H_{\theta,n}$ reemplazado por $V_{\theta,n}$, para todo $n\in\mathbb{N}$, y que también cumplen (1.11) para todo $n\in\mathbb{N}$.

La siguiente proposición muestra cómo se pueden relacionar el problema de Kiefer–Weiss y su versión modificada (cf. también la Sección 3 de Freeman and Weiss (1964)).

Proposición 1.2.2 *Sea* $(\psi^*, \phi^*) \in \mathscr{M}(\theta_*, \lambda_0, \lambda_1)$ *tal que* $\alpha = \alpha(\psi^*, \phi^*)$, $\beta = \beta(\psi^*, \phi^*)$, *y sea*

$$\Delta(\psi^*) := \sup_\theta N(\theta; \psi^*) - N(\theta_*; \psi^*). \tag{1.13}$$

Entonces se cumple

$$\sup_\theta N(\theta; \psi^*) \leq \inf_{\langle \psi, \phi \rangle \in \mathscr{S}(\alpha, \beta)} \sup_\theta N(\theta; \psi) + \Delta(\psi^*). \tag{1.14}$$

Demostración: Sea (ψ^*, ϕ^*) una prueba que satisface las condiciones de la Proposición 1.2.2, y sea (ψ, ϕ) cualquier otra prueba en $\mathscr{S}(\alpha, \beta)$. Entonces:

$$\begin{aligned} N(\theta_*; \psi^*) + \lambda_0 \alpha + \lambda_1 \beta &= N(\theta_*; \psi^*) + \lambda_0 \alpha(\psi^*, \phi^*) + \lambda_1 \beta(\psi^*, \phi^*) \\ &\leq N(\theta_*; \psi) + \lambda_0 \alpha(\psi, \phi) + \lambda_1 \beta(\psi, \phi) \\ &\leq N(\theta_*; \psi) + \lambda_0 \alpha + \lambda_1 \beta, \end{aligned}$$

donde la primera desigualdad se deduce de la Proposición 1.2.1.

Por lo tanto, se tiene que $N(\theta_*; \psi^*) \leq N(\theta_*; \psi)$, y en consecuencia:

$$\sup_\theta N(\theta; \psi^*) = N(\theta_*; \psi^*) + \Delta(\psi^*) \leq N(\theta_*; \psi) + \Delta(\psi^*) \leq \sup_\theta N(\theta; \psi) + \Delta(\psi^*),$$

lo que prueba la desigualdad (1.14). □

Obviamente, si $\Delta(\psi^*) = 0$ en la Proposición 1.2.2, entonces la prueba (ψ^*, ϕ^*) resuelve el problema original de Kiefer–Weiss con $\alpha = \alpha(\psi^*, \phi^*)$ y $\beta = \beta(\psi^*, \phi^*)$.

De este modo, la Proposición 1.2.2 permite reducir el problema original de Kiefer–Weiss a su versión modificada: basta con buscar una solución al problema modificado para algún valor de θ_* tal que $N(\theta_*; \psi^*)$ sea el máximo de $N(\theta; \psi^*)$ al variar θ.

El tratamiento del problema de Kiefer–Weiss en Weiss (1962), en particular en los casos simétricos de muestreo con distribuciones normal y Bernoulli, se basa esencialmente en esta observación.

En el caso no simétrico de observaciones Bernoulli, Freeman and Weiss (1964) proponen una elección del parámetro θ_* que hace que $\Delta(\psi^*)$ sea relativamente pequeño, lo cual permite obtener pruebas casi óptimas en el problema de Kiefer–Weiss.

Capítulo 2

El problema de Kiefer-Weiss para familias exponenciales discretas

2.1. Pruebas óptimas en el problema modificado de Kiefer-Weiss

Supongamos que las observaciones $X_1, X_2, \ldots, X_n, \ldots$ son variables aleatorias independientes e idénticamente distribuidas con una distribución perteneciente a una familia exponencial discreta uniparamétrica. Más específicamente, asumimos que

$$f_\theta(x) = \exp(\theta x - b(\theta))\, h(x), \quad x \in \mathbb{Z}^+, \tag{2.1}$$

donde $h(x)$ es una función no negativa definida en $\mathbb{Z}^+$.

Entonces, la probabilidad conjunta tiene la forma

$$f_\theta^n(x_1, \ldots, x_n) = \exp(\theta s_n - nb(\theta))\, h_n(x_1, \ldots, x_n),$$

donde $s_n = \sum_{i=1}^n x_i$ y $h_n(x_1, \ldots, x_n) = \prod_{i=1}^n h(x_i)$, para $n = 1, 2, 3, \ldots$

Es bien sabido que, bajo condiciones bastante generales, se cumple que

$$E_\theta[X_i] = b'(\theta) \quad \text{y} \quad \mathrm{Var}_\theta[X_i] = b''(\theta) > 0 \quad \text{para todo } \theta.$$

Sean $\theta_0 < \theta_1$ dos valores fijos del parámetro que definen las hipótesis H_0 y H_1.

Construyamos ahora una solución al problema modificado de Kiefer–Weiss para un valor dado de θ, cuando se realiza el muestreo a partir de una distribución del tipo (2.1).

Expresaremos las ecuaciones (1.6) – (1.8) en una forma más adecuada para su implementación computacional, utilizando la distribución del estadístico suficiente

$\sum_{i=1}^{n} X_i$ (véase Novikov et al., 2022). Esto permite evitar el manejo directo de números extremadamente pequeños que aparecen en las probabilidades conjuntas de (1.6), especialmente cuando el nivel de truncamiento H es alto.

Sea

$$g_\theta^n(s) = C_n(s)\,\exp(\theta s - nb(\theta)), \quad s \in \mathbb{Z}^+,$$

la función de masa de probabilidad del estadístico suficiente $\sum_{i=1}^{n} X_i$, para todo n natural.

Ahora definimos

$$U_{\theta,H}^H(s) = \min\left\{\lambda_0 g_{\theta_0}^H(s),\ \lambda_1 g_{\theta_1}^H(s)\right\}, \quad s \in \mathbb{Z}^+,$$

y recursivamente, para $n = H, H-1, \ldots, 2$:

$$U_{\theta,n-1}^H(s) = \min\left\{\lambda_0 g_{\theta_0}^{n-1}(s),\ \lambda_1 g_{\theta_1}^{n-1}(s),\ g_\theta^{n-1}(s) + \mathcal{J}_n U_{\theta,n}^H(s)\right\}, \quad s \in \mathbb{Z}^+,$$

donde

$$\mathcal{J}_n U(s) = \sum_x U(s+x)\, d_n(x,s), \quad s \in \mathbb{Z}^+,$$

y

$$d_n(x,s) = C_1(x)\,\frac{C_{n-1}(s)}{C_n(s+x)}, \quad x, s \in \mathbb{Z}^+.$$

Proposición 2.1.1 *Para todo $n = 1, \ldots, H$ se cumple*

$$V_{\theta,n}^H(x_1, \ldots, x_n) = \frac{U_{\theta,n}^H(s_n)\, h_n(x_1, \ldots, x_n)}{C_n(s_n)}, \tag{2.2}$$

donde $s_n = \sum_{i=1}^{n} x_i$.

Demostración: Procedemos por inducción descendente sobre $n = H, H-1, \ldots, 1$.

Del caso base $n = H$, usando la ecuación (1.6), se tiene:

$$\begin{aligned} V_{\theta,H}^H(x_1, \ldots, x_H) &= \min\left\{\lambda_0 f_{\theta_0}^H(x_1, \ldots, x_H),\ \lambda_1 f_{\theta_1}^H(x_1, \ldots, x_H)\right\} \\ &= \min\left\{\lambda_0 g_{\theta_0}^H(s_H),\ \lambda_1 g_{\theta_1}^H(s_H)\right\} \cdot \frac{h_H(x_1, \ldots, x_H)}{C_H(s_H)} \\ &= U_{\theta,H}^H(s_H)\,\frac{h_H(x_1, \ldots, x_H)}{C_H(s_H)}. \end{aligned}$$

Supongamos ahora que la igualdad (2.2) es válida para algún n con $2 \le n \le H$.

Entonces, por la ecuación (1.7) y la hipótesis de inducción:

$$\begin{aligned} V^H_{\theta,n-1}(x_1,\ldots,x_{n-1}) &= \text{mín}\left\{\lambda_0 f^{n-1}_{\theta_0},\ \lambda_1 f^{n-1}_{\theta_1},\ f^{n-1}_{\theta}\right. \\ &\qquad \left. + \sum_{x_n} U^H_{\theta,n}(s_{n-1}+x_n)\,\frac{h_n(x_1,\ldots,x_n)}{C_n(s_{n-1}+x_n)}\right\} \\ &= \text{mín}\left\{\lambda_0 g^{n-1}_{\theta_0}(s_{n-1}),\ \lambda_1 g^{n-1}_{\theta_1}(s_{n-1}),\right. \\ &\qquad \left. g^{n-1}_{\theta}(s_{n-1}) + \sum_{x_n} U^H_{\theta,n}(s_{n-1}+x_n)\,C_1(x_n)\,\frac{C_{n-1}(s_{n-1})}{C_n(s_{n-1}+x_n)}\right\} \\ &\qquad \times \frac{h_{n-1}(x_1,\ldots,x_{n-1})}{C_{n-1}(s_{n-1})} \\ &= U^H_{\theta,n-1}(s_{n-1})\,\frac{h_{n-1}(x_1,\ldots,x_{n-1})}{C_{n-1}(s_{n-1})}. \end{aligned}$$

□

Se observa fácilmente a partir de la demostración que la regla de paro óptima dada en (1.10) es equivalente a

$$\psi_n = I_{\left\{\text{mín}\left\{\lambda_0 g^n_{\theta_0},\,\lambda_1 g^n_{\theta_1}\right\} \le g^n_\theta + \mathcal{J}_{n+1} U^H_{\theta,n+1}\right\}}, \tag{2.3}$$

y que la regla de decisión (1.11) se puede escribir como

$$\phi_n = I_{\left\{\lambda_0 g^n_{\theta_0} \le \lambda_1 g^n_{\theta_1}\right\}}. \tag{2.4}$$

Así, para cualquier nivel de truncamiento H, se obtienen pruebas óptimas para el problema modificado de Kiefer–Weiss en la forma dada por (2.3)–(2.4) (o sus versiones con desigualdades estrictas), para cualquier valor fijo de θ.

Correspondientemente, la regla de paro óptima en el caso no truncado toma la forma

$$\psi_n = I_{\left\{\text{mín}\left\{\lambda_0 g^n_{\theta_0},\,\lambda_1 g^n_{\theta_1}\right\} \le g^n_\theta + \mathcal{J}_{n+1} U_{\theta,n+1}\right\}}, \tag{2.5}$$

donde $U_{\theta,n} = \lim_{H\to\infty} U^H_{\theta,n}$ para todo $n \in \mathbb{N}$.

2.2. Cotas para las regiones de continuación

En esta sección se obtienen cotas para las regiones de continuación de las pruebas óptimas en el problema modificado de Kiefer–Weiss.

Comenzamos con el caso $\theta \in (\theta_0, \theta_1)$.

De la ecuación (2.5) se deduce que $\psi_n(s) = 1$ si se cumple $\lambda_0 g^n_{\theta_0}(s) < g^n_\theta(s)$, lo cual es equivalente a

$$s > \frac{\log \lambda_0}{\theta - \theta_0} + n \cdot \frac{b(\theta) - b(\theta_0)}{\theta - \theta_0}. \tag{2.6}$$

Por lo tanto, $\psi_n(s)$ sólo puede ser igual a 0 cuando

$$s \leq B_n(\lambda_0, \theta_0, \theta),$$

donde $B_n(\lambda_0, \theta_0, \theta)$ es el mayor número entero menor o igual que el lado derecho de (2.6).

De manera análoga, se tiene que $\psi_n(s) = 1$ si $\lambda_1 g^n_{\theta_1}(s) < g^n_\theta(s)$, lo cual es equivalente a

$$s < \frac{-\log \lambda_1}{\theta_1 - \theta} + n \cdot \frac{b(\theta_1) - b(\theta)}{\theta_1 - \theta}. \tag{2.7}$$

Por consiguiente, $\psi_n(s)$ sólo puede ser igual a 0 cuando

$$s \geq A_n(\lambda_1, \theta_1, \theta),$$

donde $A_n(\lambda_1, \theta_1, \theta)$ es el menor número entero mayor o igual que el lado derecho de (2.7).

Por lo tanto,

$$\{s : \psi_n(s) < 1\} \subset \{s : A_n(\lambda_1, \theta_1, \theta) \leq s \leq B_n(\lambda_0, \theta_0, \theta)\}, \tag{2.8}$$

siempre que se cumpla

$$A_n(\lambda_1, \theta_1, \theta) \leq B_n(\lambda_0, \theta_0, \theta).$$

Entonces, el conjunto de valores de s para los cuales la prueba continúa es finito. Además, no existe ningún valor de s tal que $\psi_n(s) < 1$ (lo que implica que $\psi_n(s) \equiv 1$) si

$$A_n(\lambda_1, \theta_1, \theta) > B_n(\lambda_0, \theta_0, \theta),$$

lo cual, debido a las expresiones (2.6) y (2.7), es equivalente a

$$\frac{\log \lambda_0}{\theta - \theta_0} + \frac{\log \lambda_1}{\theta_1 - \theta} + n \left(\frac{b(\theta) - b(\theta_0)}{\theta - \theta_0} - \frac{b(\theta_1) - b(\theta)}{\theta_1 - \theta} \right) < 0. \tag{2.9}$$

Demostremos ahora que, en el caso $\theta \in (\theta_0, \theta_1)$, la expresión entre paréntesis en (2.9) es negativa.

Denotando

$$G(\theta) = (b(\theta) - b(\theta_0))(\theta_1 - \theta) - (b(\theta_1) - b(\theta))(\theta - \theta_0), \tag{2.10}$$

se observa que $G(\theta_0) = G(\theta_1) = 0$, y que su segunda derivada es

$$G''(\theta) = b''(\theta)(\theta_1 - \theta_0),$$

la cual es positiva, ya que $b''(\theta) = \mathrm{Var}_\theta(X) > 0$. Por lo tanto, $G(\theta)$ es una función convexa, y se concluye que $G(\theta) < 0$ para todo $\theta \in (\theta_0, \theta_1)$.

En consecuencia, la desigualdad (2.9) es equivalente a

$$n > \frac{\dfrac{\log \lambda_0}{\theta - \theta_0} + \dfrac{\log \lambda_1}{\theta_1 - \theta}}{\dfrac{b(\theta_1) - b(\theta)}{\theta_1 - \theta} - \dfrac{b(\theta) - b(\theta_0)}{\theta - \theta_0}}. \tag{2.11}$$

Denotemos por

$$H = H(\lambda_0, \lambda_1, \theta_0, \theta_1, \theta)$$

el mayor número entero menor o igual que el lado derecho de la desigualdad (2.11). Sabemos ahora que H representa la última etapa en la cual una prueba óptima con regla de paro ψ_n podría continuar, cuando $\theta \in (\theta_0, \theta_1)$.

No existe una cota para H si $\theta \notin (\theta_0, \theta_1)$.

Si $\theta \geq \theta_1$, entonces

$$\{s : \phi_n(s) < 1\} \subset \{s : 0 \leq s \leq B_n(\lambda_0, \theta_0, \theta)\}$$

para todo $n \in \mathbb{N}$.

Finalmente, si $\theta \leq \theta_0$, entonces

$$\{s : \phi_n(s) < 1\} \subset \{s : s \geq A_n(\lambda_1, \theta_1, \theta)\}$$

para todo $n \in \mathbb{N}$.

Para concluir esta sección, observamos que la regla de decisión óptima dada en (2.4) puede expresarse como

$$\phi_n = 1 - I_{\left[0, B_n\left(\frac{\lambda_0}{\lambda_1}, \theta_0, \theta_1\right)\right]},$$

para todo $n \in \mathbb{N}$ (cf. (2.6) y la definición de B_n).

2.3. Diseño computacional de pruebas óptimas

La base teórica se encuentra en las fórmulas (2.3) y (2.4).

Consideremos primero el caso en que $\theta \in (\theta_0, \theta_1)$.

De la expresión (2.8) se deduce que la región de continuación es finita (o vacía) para todo $n \in \mathbb{N}$. Por lo tanto, la prueba óptima queda completamente caracterizada por un conjunto de intervalos de continuación $[a_n, b_n]$, para $n = 1, \ldots, H-1$, donde H es el número máximo de observaciones que la prueba puede tomar.

Sabemos, por el apartado anterior, que $H \leq H(\lambda_0, \lambda_1, \theta_0, \theta_1, \theta)$.

Así, una prueba óptima puede obtenerse "trabajando hacia atrás" desde el valor $H = H(\lambda_0, \lambda_1, \theta_0, \theta_1, \theta)$, de la siguiente manera:

1. Establecer $n = H$.

2. Definir a_{n-1} y b_{n-1}, respectivamente, como el mínimo y el máximo valor de $s \in \mathbb{Z}^+$ para los cuales se cumple

$$g_\theta^{n-1}(s) + \mathcal{J}_n U_{\theta,n}^H(s) < \text{mín}\left\{g_{\theta_0}^{n-1}(s),\ g_{\theta_1}^{n-1}(s)\right\}. \tag{2.12}$$

   Si no existe tal valor de s, se fija $H = n-1$. Si $H = 1$, declarar el estado de "detenerse después del primer paso" y **terminar**; en caso contrario, continuar al Paso 1.

3. Para cada $s \in [a_{n-1}, b_{n-1}]$, almacenar

$$v_{n-1}(s) = \text{mín}\left\{g_\theta^{n-1}(s) + \mathcal{J}_n U_{\theta,n}^H(s),\ g_{\theta_0}^{n-1}(s),\ g_{\theta_1}^{n-1}(s)\right\}.$$

   Tener en cuenta, para su uso posterior, que

$$U_{\theta,n-1}^H(s) = \begin{cases} v_{n-1}(s), & \text{si } s \in [a_{n-1}, b_{n-1}], \\ \text{mín}\left\{g_{\theta_0}^{n-1}(s),\ g_{\theta_1}^{n-1}(s)\right\}, & \text{en caso contrario.} \end{cases}$$

4. Si $n = 2$, entonces **terminar**; de lo contrario, fijar $n = n-1$ y volver al Paso 2.

El resultado "detenerse después del primer paso" en este algoritmo significa que la prueba óptima debe finalizar tras la primera observación (por lo tanto, no es secuencial). En general, esto indica que los valores de λ_0 y/o λ_1 son demasiado pequeños como para generar una prueba secuencial significativa. En este caso, la prueba óptima se define como:

$$\psi_1 \equiv 1, \quad \phi_1 = 1 - I_{[0,\ B_1(\lambda_0/\lambda_1, \theta_0, \theta_1)]}(s).$$

Cabe señalar que, en términos generales, el valor de H no permanece fijo durante el algoritmo: d isminuye cada vez que la regla de paro óptima no encuentra una

opción para continuar hasta la etapa H. Si el algoritmo no termina en el estado "detenerse después del primer paso", se obtiene una prueba óptima caracterizada por un conjunto de intervalos de continuación $[a_n, b_n]$, para $n = 1, 2, \ldots, H-1$, donde H es el número máximo de observaciones que puede tomar la prueba.

Si $\theta \notin (\theta_0, \theta_1)$, el algoritmo anterior puede utilizarse, con algunas modificaciones, para obtener pruebas óptimas dentro de la clase de pruebas *truncadas* $\mathscr{M}_\theta^H$. Las modificaciones se refieren a la forma en que se realiza la "inducción hacia atrás" en el Paso 2, ya que, como se ha mostrado en la Subsección 2.2, cuando $\theta \notin (\theta_0, \theta_1)$ sólo existe un límite (superior o inferior) para la región de continuación. Por ejemplo, si $\theta > \theta_1$, se busca un valor de s que satisfaga la desigualdad (2.12) descendiendo desde $B_{n-1}(\lambda_0, \theta_0, \theta)$, y luego ascendiendo desde $s = 0$, obteniendo respectivamente b_{n-1} y a_{n-1} (en caso de que existan).

Si $\theta = \theta_1$, es posible restringir aún más la región de búsqueda en el Paso 2, ya que se sabe que, al muestrear a partir de una familia exponencial, la región de continuación (en caso de no ser vacía) siempre es un intervalo (véase, por ejemplo, Novikov and Popoca Jiménez, 2022). Por lo tanto, una vez que se ha determinado b_{n-1}, se puede continuar la búsqueda de a_{n-1} hacia abajo, partiendo desde dicho valor.

Existe una situación muy similar en el caso $\theta = \theta_0$, donde se busca el valor de a_{n-1} ascendiendo desde $A_{n-1}(\lambda_1, \theta_1, \theta_0)$, y luego se continúa la búsqueda hacia arriba hasta encontrar b_{n-1}.

El único caso que queda por considerar, $\theta < \theta_0$, es el más complicado, ya que no existe un límite superior para la región de continuación, a menos que la distribución de S_{n-1} esté acotada (como ocurre en el caso binomial). Por ello, en nuestra implementación computacional, se busca el valor de a_{n-1} ascendiendo desde $A_{n-1}(\lambda_1, \theta_1, \theta_0)$, y una vez encontrado, se continúa la búsqueda hacia arriba de b_{n-1} de forma indefinida. Cuando se encuentra b_{n-1}, se toma como correspondiente a la prueba óptima, salvo que exista algún valor $s > b_{n-1}$ que satisfaga la desigualdad (2.12). Lamentablemente, no existe un resultado teórico que garantice que la región de continuación sea un intervalo en este caso. En cualquier caso, el algoritmo se implementa según lo descrito, por lo que debe tenerse en cuenta que la prueba obtenida para el problema modificado de Kiefer–Weiss en el caso $\theta < \theta_0$ podría ser subóptima. Cabe mencionar que el algoritmo exacto implementado para el caso Bernoulli en Novikov et al. (2022), en los casos particulares que se analizaron experimentalmente, produjo los mismos resultados.

Como se muestra en Hawix and Schmitz (1998), las pruebas óptimas para el problema modificado de Kiefer–Weiss con $\theta > \theta_1$ o $\theta < \theta_0$ pueden no detenerse con probabilidad uno bajo las hipótesis H_0 y/o H_1, por lo que no son de especial relevancia en aplicaciones prácticas. Nuestra implementación con horizonte finito

presentada en esta sección puede ser útil en tales situaciones, ya que garantiza la terminación con probabilidad uno bajo cada una de las hipótesis, y se aproxima a la prueba óptima del problema modificado de Kiefer–Weiss cuando $H \to \infty$. No obstante, en los casos donde el número promedio de observaciones es infinito en el horizonte infinito, se deben esperar valores muy grandes de dicho número en la versión truncada. En lugar de ello, puede resultar útil incluir explícitamente el número promedio de observaciones (cuyo gran valor causa el problema) como un criterio adicional en la función de Lagrange, con un peso pequeño. Por ejemplo, si se desea minimizar $N(\theta;\psi)$ con $\theta < \theta_0$, la función de Lagrange puede incorporar adicionalmente un término $cN(\theta_1;\psi)$, con alguna constante pequeña $c > 0$. La única modificación al algoritmo descrito anteriormente consistiría en sustituir la condición (2.12) por:

$$g_\theta^{n-1}(s) + c g_{\theta_1}^{n-1}(s) + \mathcal{J}_n U_{\theta,n}^H(s) < \text{mín}\left\{g_{\theta_0}^{n-1}(s),\ g_{\theta_1}^{n-1}(s)\right\}.$$

Probablemente esto podría resultar útil en el problema aplicado mencionado por Hawix and Schmitz (1998), donde se muestra que el problema modificado de Kiefer–Weiss es, en gran medida, inaplicable. Sin embargo, en ese caso ya no se trataría de un problema de Kiefer–Weiss en sentido estricto, por lo que no entraremos en detalles. El lector interesado puede fácilmente modificar por su cuenta nuestro código para obtener una prueba significativa para ese problema. El resto de los algoritmos de evaluación de desempeño que desarrollaremos a continuación serán aplicables sin ninguna modificación.

2.4. Característica de operación, número promedio muestral y fórmulas relacionadas

En esta sección se obtienen fórmulas para calcular las probabilidades de error, el número promedio de observaciones y algunas probabilidades relacionadas, en el caso de pruebas secuenciales truncadas.

Sea $(\psi, \phi) \in \mathscr{S}^H$. Esta prueba se considerará fija a lo largo de esta subsección, por lo que omitiremos su notación en las fórmulas.

Definimos

$$a_\theta^H(s) = 1 - \phi_H(s), \quad s \in \mathbb{Z}^+, \tag{2.13}$$

y recursivamente, para $n = H-1, H-2, \ldots, 1$,

$$a_\theta^n(s) = \psi_n(s)(1-\phi_n(s)) + (1-\psi_n(s))\, E_\theta\left[a_\theta^{n+1}(s + X_{n+1})\right], \quad s \in \mathbb{Z}^+.$$

Denotemos por $D_n^H = D_n^H(\psi, \phi)$ el evento que representa "la hipótesis H_0 es aceptada en o después de la etapa n como resultado de aplicar la prueba (ψ, ϕ)".

En particular, la curva de la característica de operación se define como

$$OC_\theta(\psi, \phi) = P_\theta(D_1^H(\psi, \phi)).$$

Sea $S_n = \sum_{i=1}^n X_i$ para todo $n \in \mathbb{N}$, y $S_0 = 0$.

Proposición 2.4.1 *Para todo* $1 \leq n \leq H$ *se cumple que*

$$a_\theta^n(S_n) = P_\theta(D_n^H \mid X_1, \ldots, X_n), \tag{2.14}$$

y

$$OC_\theta(\psi, \phi) = E_\theta a_\theta^1(X_1).$$

Demostración. Por inducción hacia atrás sobre $n = H, H-1, \ldots, 1$.

Para $n = H$, se deduce de (2.13) que

$$a_\theta^H(S_H) = 1 - \phi_H(S_H) = P_\theta(D_H^H \mid X_1, \ldots, X_H),$$

(esta última igualdad se deduce de la definición de la función de decisión ϕ).

Supongamos ahora que (2.14) es válida para algún $2 \leq n \leq H$.

Entonces, se tiene:

$$\begin{aligned}
a_\theta^{n-1}(S_{n-1}) &= \psi_{n-1}(S_{n-1})(1 - \phi_{n-1}(S_{n-1})) \\
&\quad + (1 - \psi_{n-1}(S_{n-1}))E_\theta\{a_\theta^n(S_{n-1} + X_n) \mid X_1, \ldots, X_{n-1}\} \\
&= \psi_{n-1}(S_{n-1})(1 - \phi_{n-1}(S_{n-1})) \\
&\quad + (1 - \psi_{n-1}(S_{n-1}))E_\theta\{P_\theta(D_n^H \mid X_1, \ldots, X_n) \mid X_1, \ldots, X_{n-1}\} \\
&= \psi_{n-1}(S_{n-1})(1 - \phi_{n-1}(S_{n-1})) \\
&\quad + (1 - \psi_{n-1}(S_{n-1}))E_\theta\{I_{D_n^H} \mid X_1, \ldots, X_{n-1}\} \\
&= E_\theta\{\psi_{n-1}(S_{n-1})(1 - \phi_{n-1}(S_{n-1})) + (1 - \psi_{n-1}(S_{n-1}))I_{D_n^H} \mid X_1, \ldots, X_{n-1}\} \\
&= P_\theta(D_{n-1}^H \mid X_1, \ldots, X_{n-1}).
\end{aligned}$$

□

De manera similar, sea para todo $k < H$

$$b_{\theta,k}^k(s) = (1 - \psi_k(s)), \ s \in \mathbb{Z}^+, \tag{2.15}$$

y recursivamente para $n = k-1, k-2, \ldots, 1$,

$$b_{\theta,n}^k(s) = E_\theta\left[b_{\theta,n+1}^k(s + X_{n+1})\right](1 - \psi_n(s)), \quad s \in \mathbb{Z}^+. \tag{2.16}$$

Proposición 2.4.2 *Para todo* $n \leq k < H$

$$b^k_{\theta,n}(S_n) = E_\theta\{(1-\psi_n)\dots(1-\psi_k) \mid X_1,\dots,X_n\}. \tag{2.17}$$

En particular,

$$P_\theta(\tau_\psi > k) = E_\theta\, b^k_{\theta,1}(X_1).$$

Demostración. Por inducción sobre $n = k, k-1, \dots, 1$.
Para $n = k$, se deduce de (2.15) que

$$b^k_\theta(S_k) = 1 - \psi_k(S_k) = E_\theta\{1-\psi_k(S_k) \mid X_1,\dots,X_k\}.$$

Supongamos ahora que (2.17) se cumple para algún $2 \leq n \leq k$.
Entonces:

$$\begin{aligned} b^k_{\theta,n-1}(S_{n-1}) &= (1-\psi_{n-1}(S_{n-1}))\, E_\theta\{b^k_{\theta,n}(S_{n-1}+X_n) \mid X_1,\dots,X_{n-1}\} \\ &= (1-\psi_{n-1}(S_{n-1}))\, E_\theta\,\{E_\theta\{(1-\psi_n)\dots(1-\psi_k) \mid X_1,\dots,X_n\} \mid X_1,\dots,X_{n-1}\} \\ &= (1-\psi_{n-1}(S_{n-1}))\, E_\theta\{(1-\psi_n)\dots(1-\psi_k) \mid X_1,\dots,X_{n-1}\} \\ &= E_\theta\{(1-\psi_{n-1})\dots(1-\psi_k) \mid X_1,\dots,X_{n-1}\} \end{aligned}$$

□

El número promedio de observaciones puede calcularse ahora como

$$N(\theta;\psi) = \sum_{k=1}^{H} P_\theta(\tau_\psi \geq k).$$

Una forma más directa consiste en incorporar el cálculo de esta suma en (2.16). De hecho, se puede aplicar el algoritmo a cualquier regla de paro $\psi = (\psi_1, \psi_2, \dots)$ truncándola en el nivel H, es decir, definiendo

$$\psi^H = (\psi_1, \dots, \psi_H \equiv 1, \dots).$$

Corolario 2.4.3 *Sea*

$$d^H_{\theta,H-1}(s) = (1-\psi_{H-1}(s)),\ s \in \mathbb{Z}^+,$$

y, recursivamente para $n = H-2, H-3, \dots, 1$,

$$d^H_{\theta,n}(s) = (1 + E_\theta d^H_{\theta,n+1}(s+X_{n+1}))(1-\psi_n(s)),\ s \in \mathbb{Z}^+.$$

Entonces, $N(\theta;\psi^H) = 1 + E_\theta d^H_{\theta,1}(X_1)$.

Solo consideramos relevantes para aplicaciones aquellas pruebas para las cuales $N(\theta;\psi^H) \to N(\theta;\psi)$, cuando $H \to \infty$. Por lo tanto, el Corolario 2.4.3 proporciona un método para la evaluación numérica del tamaño medio de la muestra $N(\theta;\psi)$ como $\lim_{H\to\infty}(1 + E_\theta d^H_{\theta,1}(X_1))$. En particular, utilizamos este hecho para calcular el tamaño medio de la muestra de la SPRT y de las pruebas óptimas en los problemas modificados de Kiefer-Weiss cuando $\theta \notin (\theta_0, \theta_1)$ (véase el código en https://github.com/tosinabase/Kiefer-Weiss).

2.5. Resultados numéricos

El objetivo principal de este capítulo es ilustrar el uso del código en R desarrollado en ejemplos concretos basados en familias exponenciales discretas de distribuciones, para mostrar que las pruebas secuenciales obtenidas en cada uno de los ejemplos proporcionan soluciones numéricas al problema de Kiefer-Weiss, y analizar la eficiencia de dichas pruebas en comparación con la prueba secuencial de razón de probabilidades (SPRT) y con las pruebas de tamaño de muestra fijo, siempre que estas últimas tengan el mismo nivel de probabilidades de error.

El código correspondiente en R, que permite realizar todos los cálculos necesarios, está disponible en el repositorio público https://github.com/tosinabase/Kiefer-Weiss.

2.5.1. Distribución Bernoulli

Utilizamos una serie de ejemplos concretos basados en el muestreo Bernoulli presentados en Freeman and Weiss (1964), especificados por 5 pares de probabilidades de éxito hipotéticas: $\theta_0 = 0.05, \theta_1 = 0.15$; $\theta_0 = 0.1, \theta_1 = 0.2$; $\theta_0 = 0.2, \theta_1 = 0.3$; $\theta_0 = 0.4, \theta_1 = 0.5$ y $\theta_0 = 0.45, \theta_1 = 0.55$. En cada uno, empleamos un rango de probabilidades de error comúnmente utilizado en la práctica: $\alpha(= \beta) = 0.1$, 0.05, 0.025, 0.01, 0.005, 0.001 y 0.0005.

Para cada combinación de θ_0, θ_1 y $\alpha = \beta$ ejecutamos el código correspondiente a la implementación del método de la Sección 2.3, buscando una prueba con los valores de las probabilidades de error más cercanos posibles a los valores nominales. En cada caso, la diferencia relativa entre la probabilidad de error real y la nominal es menor que 0.001.

Los resultados respectivos se presentan en las Tablas 2.1 – 2.5. Para cada prueba, se incluyen en la tabla los valores correspondientes de $\theta_*, \lambda_0, \lambda_1$, así como el número promedio de observaciones $N(\theta_*;\psi^*)$ y su correspondiente $\Delta(\psi^*)$, además del cuantil 0.99 $Q_{.99}(\psi^*)$ de la distribución del tamaño de la muestra, bajo θ_*. También se

presenta el número máximo de observaciones (denotado como H en la tabla) que la prueba realmente toma.

En la segunda parte de cada tabla se presentan las características calculadas de la correspondiente prueba secuencial de razón de probabilidades (SPRT) con los valores de α y β más cercanos a los valores nominales. Estas incluyen los valores del número promedio de observaciones $N(\theta_*; W)$ y el percentil 0.99 $\mathrm{Q}_{.99}(W)$ de la distribución del tamaño de la muestra, ambos calculados en θ_*. Se calculan utilizando las fórmulas exactas de Young (1994), y no mediante las aproximaciones de Wald como en Freeman and Weiss (1964). $\log A$ y $\log B$ son los extremos del intervalo de continuación de la correspondiente SPRT.

Por último, FSS representa el valor mínimo del tamaño de muestra requerido por la prueba óptima con tamaño de muestra fijo cuyas probabilidades de error no excedan α y β. Este valor se calcula utilizando la distribución binomial de la estadística de prueba, en lugar de su aproximación normal usada en Freeman and Weiss (1964).

En la última parte de cada tabla se presentan los valores calculados de eficiencia de cada prueba con respecto a la prueba con tamaño de muestra fijo (FSST). La eficiencia se calcula como el cociente entre el valor de FSS y otras características de la prueba respectiva: $R(\psi^*) = FSS/N(\theta_*; \psi^*)$ y $QR(\psi^*) = FSS/Q_{.99}(\psi^*)$ para la prueba óptima de Kiefer-Weiss, y $R(W) = FSS/N(\theta_*; W)$ y $QR(W) = FSS/Q_{.99}(W)$ para la prueba secuencial de razón de probabilidades de Wald. Por ejemplo, $R(\psi^*) = 1.5$ significa que la prueba óptima de Kiefer-Weiss requiere, en promedio, 1.5 veces menos observaciones que la prueba correspondiente con tamaño de muestra fijo.

No hay parte correspondiente a la prueba secuencial de razón de probabilidades (SPRT) en la Tabla 2.5. La razón de esto es que, en el caso simétrico (cuando $\theta_0 = 1 - \theta_1$), a diferencia de otros casos, generalmente no es posible encontrar una SPRT que coincida, al menos aproximadamente, con los valores nominales de α y β. Esto se debe a que, en el caso simétrico, hay un número restringido de valores disponibles para $\alpha = \beta$. Por ejemplo, para $\theta_0 = 0.45$ y $\theta_1 = 0.55$, estos valores son 0.0991, 0.0826, 0.0686, 0.0568, 0.0470, etc., ninguno de los cuales coincide exactamente con los valores de $\alpha = \beta$ usados para otros pares de hipótesis. En el caso simétrico, el proceso de toma de decisiones de la SPRT es equivalente a un paseo aleatorio entre dos límites formados por dos líneas horizontales, por lo que los valores de $\alpha = \beta$ mencionados arriba se obtuvieron a partir de la fórmula bien conocida para la probabilidad de ruina del jugador (Gambler's Ruin), cuando las líneas son simétricas con respecto al eje horizontal.

Para evaluar el desempeño de las pruebas óptimas para el caso $\alpha \neq \beta$, ejecutamos el código en R sobre una cuadrícula de puntos equidistantes de λ_0 y λ_1 en la escala

logarítmica. Para cada par de θ_0 y θ_1, se encontró en cada punto de la cuadrícula la prueba óptima (ψ^*, ϕ^*) para el problema de Kiefer-Weiss. En todos los casos, se encontró que $\Delta(\psi^*) = 0$ (dentro de la precisión numérica utilizada).

Para cada par de λ_0 y λ_1, calculamos las probabilidades de error $\alpha = \alpha(\psi^*, \phi^*)$ y $\beta = \beta(\psi^*, \phi^*)$, así como los números promedio de observaciones $N(\theta_*; \psi^*)$, $N(\theta_0; \psi^*)$ y $N(\theta_1; \psi^*)$ correspondientes a la prueba óptima $(\psi^*, \phi^*) \in \mathscr{M}(\theta_*, \lambda_0, \lambda_1)$, además de una estimación del tamaño de muestra fijo requerido para alcanzar los mismos valores de α y β. En esta ocasión usamos una fórmula aproximada para el FSS basada en la aproximación normal:

$$FSS \approx \left(\frac{z_\alpha \sqrt{\theta_0(1-\theta_0)} + z_\beta \sqrt{\theta_1(1-\theta_1)}}{\theta_1 - \theta_0} \right)^2,$$

donde $z_\alpha = \Phi^{-1}(1-\alpha)$, siendo Φ la función de distribución acumulada de la distribución normal estándar. Esta forma es preferible desde el punto de vista de la suavidad en la representación gráfica que se presenta a continuación, mientras que el error relativo, en comparación con su valor exacto basado en la distribución binomial, es de hasta un 5 % para el rango de valores de α y β considerados.

En las Figuras 2.1 – 2.3, presentamos los resultados de la evaluación del desempeño para tres pares de hipótesis: $\theta_0 = 0.05$ vs. $\theta_1 = 0.15$, $\theta_0 = 0.2$ vs. $\theta_1 = 0.3$ y $\theta_0 = 0.45$ vs. $\theta_1 = 0.55$.

Cada gráfico representa una característica de desempeño (coordenada z) como función de α (coordenada x) y β (coordenada y). Para α y β, se utiliza la escala de logaritmos decimales. En cada Figura, el gráfico superior izquierdo representa la misma eficiencia utilizada en las Tablas, definida como $R(\psi^*) = FSS/N(\theta_*; \psi^*)$. El gráfico superior derecho muestra el número promedio de observaciones $N(\theta_*; \psi^*)$ como función de α y β. Los dos gráficos inferiores en cada Figura representan el desempeño de las pruebas óptimas bajo H_0 y H_1 (de forma muy similar al primer gráfico), mostrando la eficiencia definida como $R_0(\psi^*) = FSS/N(\theta_0; \psi^*)$ y $R_1(\psi^*) = FSS/N(\theta_1; \psi^*)$, respectivamente.

Los gráficos fueron preparados utilizando el paquete de visualización 3D `rgl` (Murdoch and Adler, 2021). Cada gráfico está basado en una cuadrícula de 25×25 puntos equidistantes de $\ln \lambda_0$ y $\ln \lambda_1$, ambos dentro del rango de 6 a 13.

Versiones interactivas de los gráficos están disponibles en `https://github.com/tosinabase/Kiefer-Weiss` (Novikov et al., 2021).

$\alpha = \beta$	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
θ_*	0.0768	0.0823	0.0848	0.0865	0.0874	0.0885	0.0888
λ_0	157.70	356.55	785.75	2067.24	4334.60	22957.0	46319.1
λ_1	193.35	430.27	952.99	2512.30	5183.48	27541.8	56073.6
H	128	182	237	302	345	465	508
$N(\theta_*;\psi^*)$	38.62	64.75	93.79	134.33	167.10	246.23	281.64
$\Delta(\psi^*)$	-4.E-8	1.E-4	9E-5	4.E-5	4.E-5	-3.E-6	-4.E-5
$Q_{.99}(\psi^*)$	89	134	187	244	291	397	442
$\log A$	-2.13	-2.90	-3.60	-4.54	-5.24	-6.85	-7.55
$\log B$	1.85	2.59	3.27	4.22	4.91	6.53	7.22
$N(\theta_*;W)$	40.53	71.93	109.81	173.09	228.77	388.01	469.32
$Q_{.99}(W)$	149	269	414	660	884	1502	1819
FSS	60	93	136	181	224	322	365
$\mathrm{R}(\psi^*)$	1.55	1.44	1.45	1.35	1.34	1.31	1.30
$\mathrm{QR}(\psi^*)$	0.67	0.69	0.73	0.74	0.77	0.81	0.83
$\mathrm{R}(W)$	1.48	1.29	1.24	1.05	0.98	0.83	0.78
$\mathrm{QR}(W)$	0.40	0.35	0.33	0.27	0.25	0.21	0.20

Tabla 2.1: Distribución de Bernoulli con $\theta_0 = 0.05$, $\theta_1 = 0.15$

$\alpha = \beta$	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
θ_*	0.1364	0.1394	0.1409	0.1420	0.1425	0.1432	0.1434
λ_0	246.64	557.46	1204.68	3212.66	6701.95	35404.1	72301.3
λ_1	275.38	620.37	1343.38	3595.28	7471.57	39608.9	80459.6
H	232	308	384	480	549	721	783
$N(\theta_*;\psi^*)$	56.45	95.03	138.06	198.54	246.92	364.39	416.74
$\Delta(\psi^*)$	5.E-5	9.E-7	9.E-5	5.E-7	2.E-5	2.E-5	4.E-5
$Q_{.99}(\psi^*)$	135	205	274	364	434	592	657
$\log A$	-2.13	-2.88	-3.60	-4.53	-5.24	-6.85	-7.54
$\log B$	1.95	2.70	3.42	4.35	5.04	6.66	7.35
$N(\theta_*;W)$	59.48	106.66	163.68	258.59	342.82	583.22	705.82
$Q_{.99}(W)$	224	410	636	1007	1337	2280	2759
FSS	86	135	190	272	328	479	541
$\mathrm{R}(\psi^*)$	1.52	1.42	1.38	1.37	1.33	1.31	1.30
$\mathrm{QR}(\psi^*)$	0.64	0.66	0.69	0.75	0.76	0.81	0.82
$\mathrm{R}(W)$	1.45	1.27	1.15	1.05	0.96	0.82	0.77
$\mathrm{QR}(W)$	0.38	0.33	0.30	0.27	0.25	0.21	0.20

Tabla 2.2: Distribución de Bernoulli con $\theta_0 = 0.1$, $\theta_1 = 0.2$

$\alpha = \beta$	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
θ_*	0.2435	0.2450	0.2457	0.2462	0.2464	0.2468	0.2469
λ_0	381.04	866.23	1870.27	4985.94	10346.1	54837.5	111392
λ_1	403.11	912.57	1971.23	5256.63	10880.4	57566.7	117078
H	414	539	653	786	895	1133	1242
$N(\theta_*;\psi^*)$	84.43	142.50	206.73	297.74	370.24	546.67	625.28
$\Delta(\psi^*)$	2.E-5	1.E-5	1.E-8	1.E-7	2.E-5	1.E-5	1.E-5
$Q_{.99}(\psi^*)$	204	309	410	547	649	886	989
$\log A$	-2.12	-2.88	-3.59	-4.53	-5.23	-6.84	-7.54
$\log B$	2.05	2.78	3.51	4.44	5.14	6.75	7.44
$N(\theta_*;W)$	89.41	160.30	247.69	389.73	517.48	880.57	1066.27
$Q_{.99}(W)$	346	627	968	1527	2027	3453	4181
FSS	127	204	289	402	495	713	806
$\mathrm{R}(\psi^*)$	1.50	1.43	1.40	1.35	1.34	1.30	1.29
$\mathrm{QR}(\psi^*)$	0.62	0.66	0.70	0.73	0.76	0.80	0.82
$\mathrm{R}(W)$	1.42	1.27	1.17	1.03	0.96	0.81	0.76
$\mathrm{QR}(W)$	0.37	0.33	0.30	0.26	0.24	0.21	0.19

Tabla 2.3: Distribución de Bernoulli con $\theta_0 = 0.2$, $\theta_1 = 0.3$

$\alpha = \beta$	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
θ_*	0.4490	0.4493	0.4494	0.4494	0.4495	0.4495	0.4495
λ_0	519.98	1177.67	2540.72	6804.60	14096.5	74239.7	151569
λ_1	524.39	1186.52	2561.23	6853.52	14170.5	74851.8	152375
H	575	744	893	1082	1231	1558	1698
$N(\theta_*;\psi^*)$	111.82	189.19	274.56	395.75	491.95	726.33	830.78
$\Delta(\psi^*)$	-6.E-6	-2.E-5	6.E-7	3.E-5	8.E-5	1.E-4	8.E-5
$Q_{.99}(\psi^*)$	272	409	546	726	862	1177	1313
$\log A$	-2.12	-2.86	-3.58	-4.51	-5.21	-6.82	-7.52
$\log B$	2.10	2.85	3.57	4.50	5.20	6.81	7.51
$N(\theta_*;W)$	118.79	213.16	330.10	519.05	688.86	1172.59	1420.01
$Q_{.99}(W)$	465	835	1293	2037	2703	4605	5576
FSS	168	268	384	535	655	944	1071
$\mathrm{R}(\psi^*)$	1.50	1.42	1.40	1.35	1.33	1.30	1.29
$\mathrm{QR}(\psi^*)$	0.62	0.66	0.70	0.74	0.76	0.80	0.82
$\mathrm{R}(W)$	1.41	1.26	1.16	1.03	0.95	0.81	0.75
$\mathrm{QR}(W)$	0.36	0.32	0.30	0.26	0.24	0.21	0.19

Tabla 2.4: Distribución de Bernoulli con $\theta_0 = 0.4$, $\theta_1 = 0.5$

$\alpha = \beta$	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
θ_*	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
λ_0	526.61	1193.78	2577.11	6878.97	14273.4	75384.0	153475
λ_1	526.61	1193.78	2577.11	6878.97	14273.4	75383.2	153475
H	571	733	887	1081	1227	1557	1699
$N(\theta_*;\psi^*)$	112.71	191.27	277.26	399.83	497.03	733.80	839.32
$\Delta(\psi^*)$	0.E+0	0.E+0	6.E-14	-1.E-13	-3.E-13	2.E-13	2.E-13
$Q_{.99}(\psi^*)$	274	414	551	733	871	1190	1328
FSS	163	269	383	539	661	951	1077
$\mathrm{R}(\psi^*)$	1.45	1.41	1.38	1.35	1.33	1.30	1.28
$\mathrm{QR}(\psi^*)$	0.59	0.65	0.70	0.74	0.76	0.80	0.81

Tabla 2.5: Distribución de Bernoulli con $\theta_0 = 0.45$, $\theta_1 = 0.55$

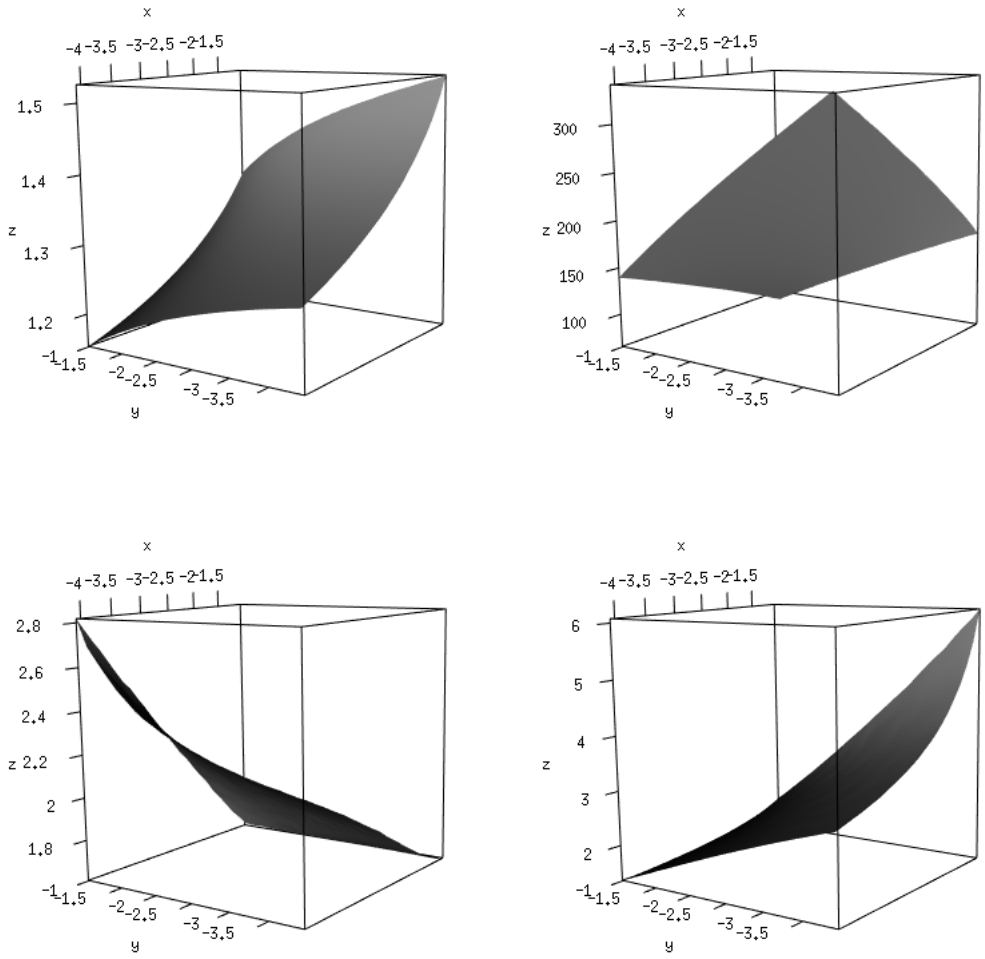


Figura 2.1: Desempeño de la prueba óptima para $\theta_0 = 0.05$ vs. $\theta_1 = 0.15$: $R(\psi^*)$ y $N(\theta_*;\psi^*)$ (arriba), $R_0(\psi^*)$ y $R_1(\psi^*)$ (abajo)

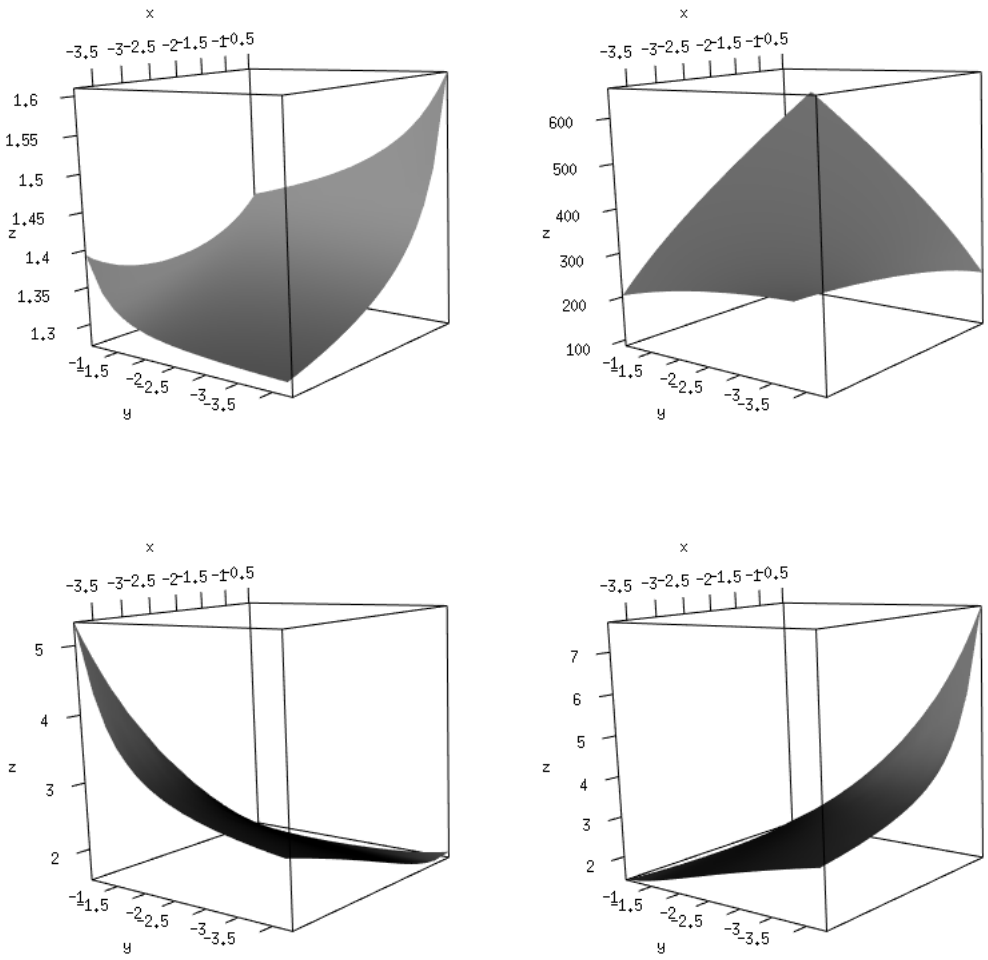


Figura 2.2: Desempeño de la prueba óptima para $\theta_0 = 0.2$ vs. $\theta_1 = 0.3$: $R(\psi^*)$ and $N(\theta_*; \psi^*)$ (arriba), $R_0(\psi^*)$ and $R_1(\psi^*)$ (abajo)

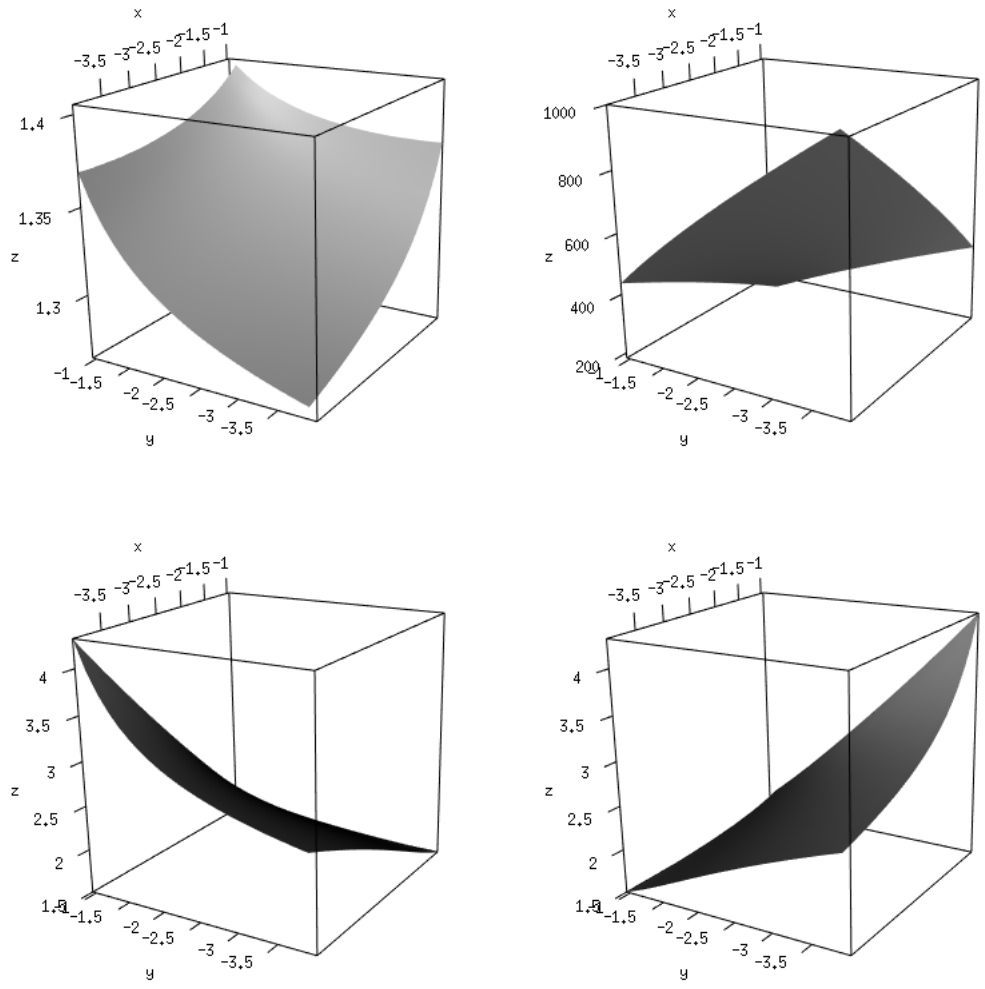


Figura 2.3: Desempeño de la prueba óptima para $\theta_0 = 0.45$ vs. $\theta_1 = 0.55$: $R(\psi^*)$ and $N(\theta_*; \psi^*)$ (arriba), $R_0(\psi^*)$ and $R_1(\psi^*)$ (abajo)

2.5.2. Familias exponenciales discretas

Utilizamos una serie de ejemplos concretos de pruebas de hipótesis para las distribuciones binomial, de Poisson y geométrica. Se prueban las siguientes hipótesis: $\theta_0 = 0.05$ vs. $\theta_1 = 0.08$ para la distribución binomial correspondiente a $n = 3$ ensayos de Bernoulli con probabilidad de éxito θ; $\theta_0 = 0.5$ vs. $\theta_1 = 0.7$ para la distribución de Poisson con media θ; y $\theta_0 = 1$ vs. $\theta_1 = 2$ para la distribución geométrica con media θ. Para cada par, empleamos una gama de probabilidades de error ampliamente utilizadas en la práctica: $\alpha(= \beta) = 0.1, 0.05, 0.025, 0.01, 0.005, 0.001$ y 0.0005.

Además, para observar el efecto de las probabilidades de error asimétricas, realizamos las mismas evaluaciones para los casos $\alpha = 0{,}1$ y $\beta = 0{,}0005$, así como para $\alpha = 0{,}0005$ y $\beta = 0{,}1$, para cada par de hipótesis.

Para cada combinación de θ_0, θ_1, α y β, resolvemos el problema correspondiente de Kiefer-Weiss siguiendo el método de la Sección 2 en Novikov et al. (2022) (utilizando la Opción 1 con la cota H definida por el lado derecho de (2.11)). Para la solución del problema modificado de Kiefer-Weiss correspondiente, empleamos los algoritmos de las Subsecciones 2.3 y 2.4.

Los únicos "valores de entrada" que se mantienen fijos en cada evaluación para cada modelo son $\theta_0, \theta_1, \alpha, \beta$. Todos los demás valores se calculan a partir de los tests óptimos del problema de Kiefer-Weiss. En particular, se busca un valor de θ que minimice

$$\Delta(\psi^*) = \sup_{\vartheta} N(\vartheta; \psi^*) - N(\theta; \psi^*),$$

con alguna elección de $(\psi^*, \phi^*) \in \mathscr{M}_\theta$, considerando todos los posibles θ. Según la Proposición 1.2.2, si $\Delta = 0$, entonces (ψ^*, ϕ^*) resuelve el problema de Kiefer-Weiss en la clase $\mathscr{S}(\alpha, \beta)$, con $\alpha = \alpha(\psi^*, \phi^*)$ y $\beta = \beta(\psi^*, \phi^*)$. En todos los ejemplos que evaluamos numéricamente, el valor aproximado de Δ resultó ser muy cercano a 0, por lo tanto, el test correspondiente (ψ^*, ϕ^*) puede considerarse una solución numérica del problema de Kiefer-Weiss. Reportamos los valores calculados de θ para cada caso evaluado del test óptimo, a fin de proporcionar el conjunto completo de parámetros que garantizan la reproducibilidad de los resultados. Además, este valor de θ corresponde al punto donde se alcanza el valor máximo del número promedio de observaciones

$$N(\theta; \psi^*) = \sup_{\vartheta} N(\vartheta; \psi^*),$$

el cual es minimizado por el test óptimo (ψ^*, ϕ^*).

El número máximo de observaciones H que toma el test óptimo (ψ^*, ϕ^*) también se determina como una característica del test óptimo (ψ^*, ϕ^*) (el algoritmo de la Subsección 2.3 realiza este trabajo).

De este modo, proporcionando los valores de θ_0, θ_1, θ, λ_0, λ_1 y H, se obtiene

una manera sencilla de resolver el problema de Kiefer-Weiss a través de la solución (ψ^*, ϕ^*) del problema modificado correspondiente.

Para cumplir con las restricciones (1.2) sobre las probabilidades de error, se buscan los multiplicadores de Lagrange λ_0 y λ_1 que proporcionen la mejor aproximación a los valores nominales de α y β. Se utiliza la minimización numérica sobre λ_0 y λ_1 buscando un valor mínimo de

$$\text{máx}\left\{\frac{|\alpha(\psi^*,\phi^*)-\alpha|}{\alpha},\ \frac{|\beta(\psi^*,\phi^*)-\beta|}{\beta}\right\}. \tag{2.18}$$

Empleamos la función `fminsearch` del paquete de R `neldermead` para el método de Nelder-Mead sin gradiente (ver Bihorel, 2022).

Una fórmula razonablemente buena para iniciar la minimización numérica de (2.18) es:

$$\lambda_0 = \kappa_0/\alpha, \quad \lambda_1 = \kappa_1/\beta,$$

donde κ_0 y κ_1 son coeficientes empíricos en el rango de 10 a 60, dependiendo del modelo.

Debido a la naturaleza discreta de las probabilidades involucradas en la evaluación, en general no hay garantía de que el mínimo de (2.18) sea igual a 0. Sin embargo, en todos los casos evaluados, las probabilidades de error reales y las nominales están aproximadamente dentro de una distancia relativa de 0.002 según (2.18), para el rango de valores de α y β considerados.

Los resultados numéricos que obtuvimos se presentan en las Tablas 2.6 a 2.9.

Para cada prueba, la tabla contiene los valores correspondientes de $\theta, \lambda_0, \lambda_1$, el número promedio de observaciones $N(\theta;(\psi^*,\phi^*))$, así como su valor correspondiente de Δ (ver (1.13)), el número promedio de observaciones bajo las dos hipótesis $N(\theta_0;(\psi^*,\phi^*))$ y $N(\theta_1;(\psi^*,\phi^*))$, y el percentil 99 $Q_{99}(\psi^*)$ de la distribución del tamaño de muestra bajo θ. También mostramos en este bloque el número máximo de observaciones (denotado por H en la tabla) que la prueba óptima realmente toma.

El segundo bloque (inferior) de cada tabla muestra las características calculadas de la correspondiente "prueba secuencial de razón de probabilidades" (SPRT) con los valores de α y β más cercanos a los valores nominales. Se ha utilizado el mismo procedimiento de minimización numérica mencionado anteriormente, esta vez sobre los límites de continuación A y B de la SPRT. Colocamos en la tabla los valores del número promedio de observaciones de la SPRT de Wald, $N(\theta;W)$, junto con los números promedio de observaciones bajo ambas hipótesis, $N(\theta_0;W)$ y $N(\theta_1;W)$, y el percentil 99 $Q_{99}(W)$ de la distribución del tamaño de muestra, calculado bajo θ. Todas las características se calculan usando las fórmulas exactas de la Subsección 2.4, con un horizonte suficientemente grande H. Los valores $\log(A)$ y $\log(B)$, que son los

extremos del intervalo de continuación de la SPRT correspondiente, se proporcionan para garantizar la reproducibilidad de los resultados.

Esto permite una forma sencilla de comparar las características respectivas de las dos pruebas óptimas: la de Kiefer-Weiss frente a la SPRT de Wald. Por ejemplo, se puede observar cuánto pierde en promedio la prueba óptima de Kiefer-Weiss frente a la SPRT bajo H_0 o H_1, donde la SPRT es óptima (es decir, comparar $N(\theta_0;\psi^*)$ con $N(\theta_0;W)$, y/o $N(\theta_1;\psi^*)$ con $N(\theta_1;W)$); o comparar el número promedio máximo de observaciones que alcanza la prueba óptima de Kiefer-Weiss, $N(\theta;\psi^*) = \sup_\vartheta N(\vartheta;\psi^*)$, con el número promedio de observaciones $N(\theta;W)$ de la SPRT en el mismo punto del parámetro. La comparación de los percentiles $Q_{99}(\theta;\psi^*)$ y $Q_{99}(\theta;W)$ puede ser aún más reveladora, ya que es bien sabido que la SPRT tiende a requerir más observaciones cuando el verdadero valor del parámetro θ está entre θ_0 y θ_1.

Finalmente, FSS representa el valor mínimo del tamaño de muestra requerido por la prueba óptima con tamaño de muestra fijo que cumple con las probabilidades de error α y β. Para valores dados de α y β, FSS se calcula como

$$n^* + \frac{\beta(n^*) - \beta}{\beta(n^*) - \beta(n^*+1)},$$

donde $n^* = n^*(\alpha,\beta)$ es el mayor entero n tal que la probabilidad de error de tipo II, $\beta(n)$, de la prueba más poderosa de nivel α (de Neyman-Pearson) es mayor o igual que β.

En la última parte de cada tabla (el bloque más inferior), se muestran los valores calculados de eficiencia de cada prueba con respecto a las pruebas FSS. La eficiencia se calcula como el cociente entre el valor de FSS y otras características de la prueba respectiva:

$$R(\psi^*) = \frac{\text{FSS}}{N(\theta;\psi^*)}$$

$$R_0(\psi^*) = \frac{\text{FSS}}{N(\theta_0;\psi^*)},$$

$$R_1(\psi^*) = \frac{\text{FSS}}{N(\theta_1;\psi^*)},$$

$$QR(\psi^*) = \frac{\text{FSS}}{Q_{99}(\psi^*)}$$

para la prueba óptima de Kiefer-Weiss, y

$$R(W) = \frac{\text{FSS}}{N(\theta;W)},$$

$\alpha = \beta$	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
λ_0	305.94	691.65	1495.23	3993.13	8275.87	43707.77	88888.07
λ_1	326.39	737.05	1596.43	4262.75	8815.25	46564.03	94690.75
θ	0.58464	0.58794	0.58953	0.59072	0.59130	0.59206	0.59225
H	353	442	533	649	745	935	1024
$N(\theta;\psi^*)$	67.93	114.80	166.46	239.77	297.96	439.60	502.73
$N(\theta_0;\psi^*)$	57.06	87.92	117.97	156.92	185.92	252.90	281.79
$N(\theta_1;\psi^*)$	51.66	79.55	106.80	141.98	168.15	228.59	254.63
$Q_{99}(\psi^*)$	165	247	331	440	522	712	794
Δ	1E-06	2E-06	2E-05	9E-06	3E-06	8E-06	4E-05
log(A)	-0.916	-1.240	-1.553	-1.957	-2.260	-2.961	-3.262
log(B)	0.868	1.191	1.504	1.908	2.212	2.912	3.214
$N(\theta;W)$	72.28	129.16	199.74	313.98	416.52	708.62	857.74
$N(\theta_0;W)$	55.55	83.53	109.64	141.86	165.01	217.05	239.08
$N(\theta_1;W)$	50.06	75.11	98.48	127.26	148.04	194.56	214.22
$Q_{99}(W)$	281	504	780	1229	1632	2779	3364
FSS	98.07	161.05	229.01	322.43	395.11	568.75	644.84
$R(\psi^*)$	1.44	1.40	1.38	1.34	1.33	1.29	1.28
$R_0(\psi^*)$	1.72	1.83	1.94	2.05	2.13	2.25	2.29
$R_1(\psi^*)$	1.90	2.02	2.14	2.27	2.35	2.49	2.53
$QR(\psi^*)$	0.59	0.65	0.69	0.73	0.76	0.80	0.81
$R(W)$	1.36	1.25	1.15	1.03	0.95	0.80	0.75
$R_0(W)$	1.77	1.93	2.09	2.27	2.39	2.62	2.70
$R_1(W)$	1.96	2.14	2.33	2.53	2.67	2.92	3.01
$QR(W)$	0.35	0.32	0.29	0.26	0.24	0.20	0.19

Tabla 2.6: Distribución de Poisson con media $\theta_0 = 0.5$ vs. $\theta_1 = 0.7$

$$R_0(W) = \frac{\text{FSS}}{N(\theta_0; W)},$$

$$R_1(W) = \frac{\text{FSS}}{N(\theta_1; W)},$$

$$QR(W) = \frac{\text{FSS}}{Q_{99}(W)}$$

para la prueba secuencial de razón de probabilidades de Wald. Por ejemplo, $R_0(\psi^*) = 2$ significa que la prueba óptima de Kiefer-Weiss requiere, en promedio, 2 veces menos observaciones bajo H_0 que la prueba correspondiente con tamaño de muestra fijo.

$\alpha = \beta$	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
λ_0	69.00	154.09	333.06	893.28	1848.87	9762.77	19882.77
λ_1	84.38	189.88	408.32	1092.24	2258.64	11940.38	24318.97
θ	1.27794	1.31841	1.33953	1.35556	1.36336	1.37428	1.37741
H	74	96	118	145	167	213	233
$N(\theta;\psi^*)$	17.08	28.37	40.81	58.44	72.52	106.58	121.76
$N(\theta_0;\psi^*)$	15.43	23.60	31.62	41.92	49.66	67.51	75.212
$N(\theta_1;\psi^*)$	11.63	17.53	23.35	30.75	36.32	49.11	54.57
$Q_{99}(\psi^*)$	41	60	80	106	126	172	192
Δ	3E-07	1E-08	1E-06	9E-09	8E-08	-2E-07	-5E-08
$\log A$	-0.8920	-1.2098	-1.5170	-1.9213	-2.2260	-2.9269	-3.2278
$\log B$	0.7318	1.0452	1.3615	1.7649	2.0668	2.7692	3.0698
$N(\theta;W)$	18.24	31.59	48.32	75.44	99.78	168.99	204.24
$N(\theta_0;W)$	15.30	22.62	29.64	38.30	44.53	58.60	64.53
$N(\theta_1;W)$	11.42	16.61	21.48	27.51	31.87	41.60	45.71
$Q_{99}(W)$	67	119	184	290	384	654	791
FSS	23.83	39.00	55.29	77.74	95.21	136.89	155.21
$R(\psi^*)$	1.40	1.37	1.35	1.33	1.31	1.28	1.27
$R_0(\psi^*)$	1.54	1.65	1.75	1.85	1.92	2.03	2.06
$R_1(\psi^*)$	2.05	2.22	2.37	2.53	2.62	2.79	2.84
$QR(\psi^*)$	0.58	0.65	0.69	0.73	0.76	0.80	0.81
$R(W)$	1.31	1.23	1.14	1.03	0.95	0.81	0.76
$R_0(W)$	1.56	1.72	1.87	2.03	2.14	2.34	2.41
$R_1(W)$	2.09	2.35	2.57	2.83	2.99	3.29	3.40
$QR(W)$	0.36	0.33	0.30	0.27	0.25	0.21	0.20

Tabla 2.7: Distribución geométrica con media $\theta_0 = 1$ vs. $\theta_1 = 2$

2.6. Análisis de los resultados numéricos

2.6.1. Distribución Bernoulli

Las Tablas 2.1 – 2.5 son útiles para analizar la eficiencia relativa $R(\psi^*)$ de las pruebas óptimas con respecto a la prueba de tamaño de muestra fijo. Se observa claramente que no varía mucho en todo el rango de valores de α y β calculados. Su valor más bajo es aproximadamente 1.3 y se alcanza en los valores mínimos de α y β analizados. Existe una clara tendencia a que la eficiencia relativa disminuya cuando $\alpha = \beta \to 0$. Sería interesante contar con una cota teórica inferior significativa para su valor límite, pero dudamos que alguna vez pueda obtenerse. Lo único que sabemos es que, por definición, la eficiencia relativa no puede ser menor que 1.

La eficiencia relativa $QR(\psi^*)$ basada en el percentil 0.99 de la prueba óptima de Kiefer-Weiss se comporta bastante bien, manteniéndose en un nivel aproximado entre 0.6 y 0.8 para valores prácticos de $\alpha = \beta$.

$\alpha = \beta$	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
λ_0	450.00	1020.19	2201.68	5880.48	12177.81	64446.61	131326.7
λ_1	489.75	1110.18	2393.14	6398.75	13250.02	70018.20	142274.9
θ	0.06193	0.06263	0.06296	0.06316	0.06328	0.06345	0.06350
H	484	630	771	943	1068	1355	1480
$N(\theta;\psi^*)$	101.13	171.00	247.90	357.06	443.72	654.74	748.81
$N(\theta_0;\psi^*)$	86.28	133.00	178.58	237.31	281.04	382.50	426.12
$N(\theta_1;\psi^*)$	75.46	116.41	156.41	208.08	246.49	335.21	373.45
$Q_{99}(\psi^*)$	247	369	493	654	778	1060	1184
Δ	-3E-07	-1E-06	2E-04	2E-04	4E-04	-7E-08	5E-05
log(A)	-2.1517	-2.8990	-3.6164	-4.5494	-5.2469	-6.8607	-7.5540
log(B)	2.0034	2.7527	3.4713	4.4033	5.1012	6.7147	7.4086
$N(\theta;W)$	107.24	192.16	296.92	466.76	618.96	1053.14	1275.06
$N(\theta_0;W)$	83.91	126.52	166.12	215.01	250.13	329.02	362.47
$N(\theta_0;W)$	73.07	109.94	144.16	186.36	216.82	285.00	313.86
$Q_{99}(W)$	414	748	1159	1826	2423	4124	4997
FSS	146.62	240.94	341.25	480.98	589.32	847.40	961.23
$R(\psi^*)$	1.45	1.41	1.38	1.35	1.33	1.29	1.28
$R_0(\psi^*)$	1.70	1.81	1.91	2.03	2.10	2.22	2.26
$R_1(\psi^*)$	1.94	2.07	2.18	2.31	2.39	2.53	2.57
$QR(\psi^*)$	0.59	0.65	0.69	0.74	0.76	0.80	0.81
$R(W)$	1.37	1.25	1.15	1.03	0.95	0.80	0.75
$R_0(W)$	1.75	1.90	2.05	2.24	2.36	2.58	2.65
$R_1(W)$	2.01	2.19	2.37	2.58	2.72	2.97	3.06
$QR(W)$	0.35	0.32	0.29	0.26	0.24	0.21	0.19

Tabla 2.8: Distribución binomial con $\theta_0 = 0.05$ vs. $\theta_1 = 0.08$ y $n = 3$

La eficiencia relativa de la prueba secuencial de razón de probabilidades (SPRT), basada en el número promedio de observaciones evaluado en θ_*, cae a aproximadamente $0.7 - 0.8$ para niveles bajos de α y β, lo cual aún podría considerarse tolerable, al menos comparable con la eficiencia 1.3 de la prueba óptima de Kiefer-Weiss. Lo que resulta catastrófico para la SPRT es su baja eficiencia $QR(W)$, basada en el percentil 0.99 de la distribución del tamaño de la muestra bajo θ_*, que llega a ser tan baja como aproximadamente $0.3 - 0.2$, lo que significa que el percentil 0.99 puede alcanzar un nivel de 3 a 5 veces superior al del tamaño de muestra fijo.

La eficiencia relativa basada en el número promedio de observaciones calculado en θ_* es importante como una guía para la construcción de pruebas óptimas, pero tiene una importancia limitada para las aplicaciones prácticas: la función de característica de operación evaluada en θ^* es siempre cercana a 0.5; por lo tanto, en esta circunstancia, la prueba óptima es tan útil como lanzar una moneda. El criterio "minimax" de Kiefer-Weiss se asemeja más a un principio de "robustez" para la construcción de pruebas. Una discusión más profunda sobre este aspecto del

	$\alpha = 0.1, \beta = 0.0005$			$\alpha = 0.0005, \beta = 0.1$		
Modelo	Binomial	Poisson	Geométrica	Binomial	Poisson	Geométrica
(θ_0,θ_1)	(0.05,0.08)	(0.5,0.7)	(1.0,2.0)	(0.05,0.08)	(0.5,0.7)	(1.0,2.0)
λ_0	948.57	640.92	150.43	80909.41	54753.29	11788.63
λ_1	91786.79	60360.17	15823.13	978.27	656.25	161.15
θ	0.05551	0.53918	1.12992	0.07145	0.64532	1.62774
H	1172	806	200	1058	741	151
$N(\theta;\psi^*)$	350.27	232.95	60.47	324.14	219.70	50.07
$N(\theta_0;\psi^*)$	320.30	211.46	56.82	135.48	89.49	24.16
$N(\theta_1;\psi^*)$	116.76	79.88	17.36	279.23	190.56	40.83
$Q_{99}(\psi^*)$	695	464	118	655	443	102
Δ	-8E-05	6E-06	-5E-08	6E-05	3E-05	-4E-09
log(A)	-7.4481	-3.2167	-3.1829	-2.2541[1]	-0.9620	-0.9276
log(B)	2.1115	0.9132	0.7673	7.1464[1]	3.1680	3.0242
$N(\theta;W)$	392.78	263.20	65.89	373.89[1]	258.21	61.57
$N(\theta_0;W)$	311.04	205.19	55.50	109.62[1]	72.37	19.56
$N(\theta_1;W)$	95.35	65.17	14.54	263.52[1]	183.89	39.27
$Q_{99}(W)$	1293	879	200	1372[1]	937	241
FSS	478.61	319.13	81.10	449.66	303.78	69.14
$R(\psi^*)$	1.37	1.37	1.34	1.39	1.38	1.38
$R_0(\psi^*)$	1.49	1.51	1.43	3.32	3.39	2.86
$R_1(\psi^*)$	4.10	4.00	4.67	1.61	1.59	1.69
$QR(\psi^*)$	0.69	0.69	0.69	0.69	0.69	0.68
$R(W)$	1.22	1.21	1.23	1.20[1]	1.18	1.12
$R_0(W)$	1.54	1.56	1.46	4.10[1]	4.20	3.53
$R_1(W)$	5.02	4.90	5.58	1.71[1]	1.65	1.76
$QR(W)$	0.37	0.36	0.41	0.38[1]	0.32	0.29

[1] Esta columna corresponde al valor real de $\alpha = 0{,}000585$, no se pudo lograr una mejor aproximación para $\alpha = 0{,}0005$

Tabla 2.9: Comparación de pruebas para probabilidades de error asimétricas

problema de Kiefer-Weiss y cuestiones relacionadas puede encontrarse en Fauß and Poor (2020).

Lo que realmente importa para el desempeño de una prueba es su comportamiento bajo H_0 y H_1. Para analizar esto, definimos dos características adicionales: $R_0(\psi^*) = FSS/N(\theta_0;\psi^*)$ y $R_1(\psi^*) = FSS/N(\theta_1;\psi^*)$. El comportamiento de las tres características, R, R_0 y R_1, puede observarse en las Figuras 2.1 – 2.3 como función de α y β (en escala logarítmica).

Se observa claramente que $R(\psi^*)$ mantiene su nivel en aproximadamente 1.3 – 1.5, ahora en todo el rango de α y β, no solo cuando son iguales.

La eficiencia relativa bajo H_0, R_0, tiende a tener valores más altos en casos muy asimétricos con α pequeña y β grande, y valores más bajos cuando α es grande y β pequeña, situándose entre 1.8 y 2.4 en las cercanías de la diagonal $\alpha = \beta$. R_1 se

comporta casi simétricamente a R_0 con respecto a la diagonal $\alpha = \beta$. En el caso simétrico $\theta_0 = 1 - \theta_1$, los valores de R_0 y R_1 son exactamente simétricos.

En general, las pruebas óptimas en el contexto de Kiefer-Weiss mantienen buenos niveles de eficiencia, requiriendo, bajo H_0 y H_1, en promedio entre 1.8 y 2.4 veces menos observaciones que la mejor prueba con tamaño fijo (para $\alpha \approx \beta$), y entre 1.3 y 1.5 veces menos en el peor de los casos, cuando la hipótesis está mal especificada.

2.6.2. Familias exponenciales discretas

Las Tablas 2.6 – 2.9 son convenientes para analizar la eficiencia relativa de las pruebas óptimas con respecto a la prueba de tamaño de muestra fijo (FSS).

Se observa claramente que las tres eficiencias $R(\psi^*)$, $R_0(\psi^*)$ y $R_1(\psi^*)$ no varían mucho en todo el rango de valores de α y β evaluados.

El valor más bajo de $R(\psi^*)$ está ligeramente por debajo de 1.3 y se alcanza en los valores mínimos de $\alpha = \beta$ considerados. Hay una tendencia clara de disminución de la eficiencia relativa a medida que $\alpha = \beta \to 0$. Como referencia, se puede tener en cuenta que la eficiencia relativa $R(\psi^*)$, por definición, no puede caer por debajo de 1. Se observa claramente en la Tabla 2.9 que $R(\psi^*)$ mantiene su nivel alrededor de 1.4 incluso para valores muy asimétricos de α y β.

Tanto $R_0(\psi^*)$ como $R_1(\psi^*)$ varían entre aproximadamente 1.8 y 2.8 para $\alpha = \beta$, alcanzando sus valores máximos para valores pequeños de $\alpha = \beta$.

La eficiencia relativa bajo H_1, $R_1(\psi^*)$, tiende a tener valores más altos (hasta aproximadamente 4.0 – 4.5) en el caso muy asimétrico de $\alpha = 0.1$ y $\beta = 0.0005$, mientras que $R_0(\psi^*)$ se mantiene en aproximadamente 1.5, lo cual es ligeramente menor que en el caso de α y β iguales. El comportamiento de $R_0(\psi^*)$ y $R_1(\psi^*)$ para el otro par de valores extremos (cuando $\alpha = 0.0005$ y $\beta = 0.1$) está casi invertido.

La eficiencia relativa $QR(\psi^*)$ basada en el percentil 99 de la prueba óptima de Kiefer-Weiss se comporta bastante bien, manteniéndose en un nivel aproximado de 0.6 a 0.8 para todos los niveles de α y β evaluados.

La eficiencia relativa de la prueba secuencial de razón de probabilidades (SPRT), $R(\theta; W)$, baja a aproximadamente 0.7 – 0.8 para niveles bajos de α y β, lo cual sigue siendo comparable con la eficiencia de aproximadamente 1.3 de la prueba óptima de Kiefer-Weiss. Pero la SPRT muestra una eficiencia muy baja basada en cuantiles, $QR(W)$, que puede ser tan baja como aproximadamente 0.3 a 0.2, lo que significa que su tamaño de muestra con 1 % de probabilidad puede ser 3 – 5 veces mayor que el tamaño fijo de muestra.

En general, resulta notable que el patrón de la eficiencia relativa sea en gran medida el mismo tanto entre los tres modelos que consideramos en este trabajo, como entre estos y el modelo Bernoulli de 2.6.1 (Novikov et al., 2022). Al comparar

los valores de eficiencia relativa $R(\psi^*)$ para el mismo nivel $\alpha = \beta$ entre los cuatro modelos, se observan resultados casi idénticos. $R_0(\psi^*)$ y $R_1(\psi^*)$ muestran un comportamiento muy similar.

Lo mismo ocurre con las eficiencias calculadas para la prueba secuencial de razón de probabilidades (SPRT) (al comparar $R(W)$, $R_0(W)$ o $R_1(W)$ entre los modelos para el mismo nivel de $\alpha = \beta$). Incluso las eficiencias basadas en cuantiles $QR(\psi^*)$ son prácticamente idénticas entre los modelos, y lo mismo ocurre con $QR(W)$.

Aproximadamente lo mismo sucede en ambos casos de probabilidades de error asimétricas (véase la Tabla 2.9):

Capítulo 3

El problema de Kiefer-Weiss para familias exponenciales continuas

3.1. Solución del problema de Kiefer-Weiss modificado para familias exponenciales

Sean las observaciones $X_1, X_2, \ldots, X_n, \ldots$ variables aleatorias independientes e idénticamente distribuidas siguiendo una distribución de una familia exponencial uniparamétrica en la recta real $\mathbb{R}$. Más específicamente, supongamos que la distribución de X_i es absolutamente continua con respecto a la medida de Lebesgue en $\mathbb{R}$, y su función de densidad de probabilidad (derivada de Radon-Nikodym) tiene la forma

$$f_\theta(x) = \exp(\theta x - b(\theta))h(x),\ x \in \mathbb{R},\ \theta \in \Theta, \tag{3.1}$$

donde $h(x)$ es una función medible no negativa en $\mathbb{R}$, y

$$\Theta = \{\theta \in \mathbb{R} : \int_{\mathbb{R}} \exp(\theta x)h(x)dx < \infty\}.$$

Suponemos que los primeros dos momentos de X_i son finitos, entonces $b'(\theta) = E_\theta X_i$ y $b''(\theta) = Var_\theta X_i > 0$, $\theta \in \Theta$.

Se pueden considerar familias exponenciales más generales usando reparametrizaciones. Todos nuestros resultados también son aplicables para el caso general, bajo supuestos naturales sobre la reparametrización.

Bajo la suposición (3.1), las pruebas óptimas en el problema modificado de Kiefer–Weiss son funciones de la estadística suficiente $S_n = \sum_{i=1}^n X_i$. Expresemos las pruebas óptimas en estos términos.

Para un vector de muestra $X_1, X_2, \ldots, X_n$, la función de densidad conjunta es

$$f_\theta^n(x_1, \ldots, x_n) = \exp(\theta s_n - nb(\theta))h_n(x_1, \ldots, x_n), \quad (x_1, \ldots, x_n) \in \mathbb{R}^n,$$

donde $s_n = \sum_{i=1}^n x_i$, y $h_n(x_1, \ldots, x_n) = \prod_{i=1}^n h(x_i)$, $n = 1, 2, \ldots$

Definimos

$$g_\theta^n(s) = \exp(\theta s - nb(\theta)), \ s \in \mathbb{R}.$$

Sean $\theta_0 \neq \theta_1$, $\theta_0, \theta_1 \in \Theta$ dos valores de parámetros fijos que identifican las hipótesis H_0 y H_1. Entonces, a partir de

$$U_{\theta,H}^H(s) = \text{mín}\{\lambda_0 g_{\theta_0}^H(s), \lambda_1 g_{\theta_1}^H(s)\}, \ s \in \mathbb{R}, \tag{3.2}$$

definimos recursivamente sobre $n = H, \ldots, 2$

$$U_{\theta,n-1}^H(s) = \text{mín}\{\lambda_0 g_{\theta_0}^{n-1}(s), \lambda_1 g_{\theta_1}^{n-1}(s), g_\theta^{n-1}(s) + \mathcal{J}U_{\theta,n}^H(s)\}, \ s \in \mathbb{R}, \tag{3.3}$$

donde

$$\mathcal{J}U(s) = \int_{\mathbb{R}} U(s+x)h(x)dx, \ s \in \mathbb{R}. \tag{3.4}$$

La integral en (3.4) es con respecto a la medida de Lebesgue en $\mathbb{R}$.

Entonces las fórmulas de inducción hacia atrás (1.6)–(1.8) son equivalentes a (3.2)–(3.4), como se ve en la siguiente proposición.

Proposición 3.1.1

$$V_{\theta,n}^H(x_1, \ldots, x_n) = U_{\theta,n}^H(s_n)h_n(x_1, \ldots, x_n), \tag{3.5}$$

for $n = 1, \ldots, H$.

Demostración. Por inducción sobre $n = H, H-1, \ldots, 1$.

De (1.6) se deduce que

$$\begin{aligned} V_{\theta,H}^H(x_1, \ldots, x_H) &= \text{mín}\left\{\lambda_0 f_{\theta_0}^H(x_1, \ldots, x_H), \ \lambda_1 f_{\theta_1}^H(x_1, \ldots, x_H)\right\} \\ &= \text{mín}\left\{\lambda_0 g_{\theta_0}^H(s_H), \ \lambda_1 g_{\theta_1}^H(s_H)\right\} h_H(x_1, \ldots, x_H) \\ &= U_{\theta,H}^H(s_H)\, h_H(x_1, \ldots, x_H). \end{aligned}$$

Supongamos ahora que (3.5) se cumple para algún $2 \leq n \leq H$. Entonces, en

virtud de (1.7),

$$\begin{aligned} V^H_{\theta,n-1} &= \text{mín}\left\{\lambda_0 f^{n-1}_{\theta_0},\ \lambda_1 f^{n-1}_{\theta_1},\ f^{n-1}_{\theta} + \int_{\mathbb{R}} U^H_{\theta,n}(x_1+\cdots+x_n)h_n(x_1,\ldots,x_n)\,dx_n\right\} \\ &= \text{mín}\left\{\lambda_0 g^{n-1}_{\theta_0},\ \lambda_1 g^{n-1}_{\theta_1},\ g^{n-1}_{\theta} + \int_{\mathbb{R}} U^H_{\theta,n}(s_{n-1}+x_n)h(x_n)\,dx_n\right\} \\ &\quad \times h_{n-1}(x_1,\ldots,x_{n-1}) \\ &= U^H_{\theta,n-1}(s_{n-1})\,h_{n-1}(x_1,\ldots,x_{n-1}). \end{aligned}$$

□

Ahora se ve fácilmente que la regla de paro óptima (1.10) es equivalente a

$$\psi_n = I_{\{\text{mín}\{\lambda_0 g^n_{\theta_0},\lambda_1 g^n_{\theta_1}\}\le g^n_\theta + \mathcal{J}U^H_{\theta,n+1}\}}, \tag{3.6}$$

y regla de decisión (1.11) a

$$\phi_n = I_{\{\lambda_0 g^n_{\theta_0}\le\lambda_1 g^n_{\theta_1}\}}. \tag{3.7}$$

Ahora, para cualquier nivel de truncamiento H, disponemos de pruebas óptimas $\mathscr{M}^H_\theta$ para el problema modificado de Kiefer–Weiss en la forma (3.6) – (3.7), para cualquier valor fijo de θ. En consecuencia, la expresión (1.12) adquiere la forma

$$\psi_n = I_{\left\{\text{mín}\{\lambda_0 g^n_{\theta_0},\lambda_1 g^n_{\theta_1}\}\le g^n_\theta + \mathcal{J}U_{\theta,n+1}\right\}},$$

donde $U_{\theta,n} = \lim_{H\to\infty} U^H_{\theta,n}$, para todo n natural.

Se observa a partir de (3.6) y (3.7) que las pruebas óptimas en los problemas de Kiefer–Weiss pueden expresarse únicamente en función de la estadística suficiente S_n cuando se muestrea de una familia exponencial de distribuciones.

En términos generales, no existe una forma cerrada para las funciones $U^H_{\theta,n}$ de las cuales dependen las pruebas óptimas, por lo que proponemos una evaluación numérica utilizando cuadraturas para el integral en (3.4), dentro del procedimiento de inducción hacia atrás dado por (3.3).

3.2. Algoritmos para el cálculo de planes de muestreo óptimos

En esta sección describimos algoritmos computacionales para el problema modificado de Kiefer–Weiss en el caso en que las distribuciones hipotéticas provienen de una familia exponencial de tipo continuo. La aplicación al problema de Kiefer–Weiss se deduce de manera directa a partir de la Proposición 1.2.1 (véase también la Pro-

posición 2 en Novikov et al., 2022).

El código correspondiente en R, en el que se implementan las fórmulas presentadas aquí, está disponible en el repositorio público https://github.com/tosinabase/Kiefer-Weiss.

3.2.1. Cotas para las regiones de continuación

En esta parte, obtenemos cotas para las regiones de continuación de las pruebas óptimas en el problema modificado de Kiefer–Weiss y determinamos una cota superior para el número máximo de observaciones requeridas por las pruebas óptimas, cuando $\theta \in (\theta_0, \theta_1)$.

Sea (ψ, ϕ) igual a $\mathscr{M}_\theta$ o bien a $\mathscr{M}_\theta^H$. Se deduce de (2.5) (o de (2.3)) que $\psi_n(s) = 1$ si $\lambda_0 g_{\theta_0}^n(s) < g_\theta^n(s)$; y esto es equivalente a

$$s > \frac{\log \lambda_0}{\theta - \theta_0} + n\frac{b(\theta) - b(\theta_0)}{\theta - \theta_0} = B_n(\lambda_0, \theta_0, \theta). \tag{3.8}$$

Por lo tanto, $\psi_n(s)$ puede ser estrictamente menor que 1 (lo que significa continuación) solamente cuando $s \leq B_n(\lambda_0, \theta_0, \theta)$.

De manera análoga, $\psi_n(s) = 1$ si $\lambda_1 g_{\theta_1}^n(s) < g_\theta^n(s)$, lo cual es equivalente a

$$s < \frac{-\log \lambda_1}{\theta_1 - \theta} + n\frac{b(\theta_1) - b(\theta)}{\theta_1 - \theta} = B_n(1/\lambda_1, \theta, \theta_1), \tag{3.9}$$

por lo tanto, $\psi_n(s)$ puede ser estrictamente menor que 1 solamente cuando $s \geq B_n(1/\lambda_1, \theta, \theta_1)$.

Por lo tanto,

$$\{s : \psi_n(s) < 1\} \subset [B_n(1/\lambda_1, \theta, \theta_1),\ B_n(\lambda_0, \theta_0, \theta)]\,; \tag{3.10}$$

así, hemos encontrado cotas para la región de continuación (posiblemente vacía) en cualquier etapa del experimento, válidas tanto para $\mathscr{M}_\theta$ como para $\mathscr{M}_\theta^H$.

Lorden (1980) demostró que la región de continuación, en cualquier etapa, es un intervalo (si no es vacía), cuando θ está estrictamente entre θ_0 y θ_1.

De (3.10) se deduce que no existe ningún s tal que $\psi_n(s) < 1$ (lo que implica que $\psi_n(s) \equiv 1$) si

$$B_n(1/\lambda_1, \theta, \theta_1) > B_n(\lambda_0, \theta_0, \theta),$$

lo cual es equivalente, en virtud de (3.8) y (3.9), a

$$\frac{\log \lambda_0}{\theta - \theta_0} + \frac{\log \lambda_1}{\theta_1 - \theta} + n\left(\frac{b(\theta) - b(\theta_0)}{\theta - \theta_0} - \frac{b(\theta_1) - b(\theta)}{\theta_1 - \theta}\right) < 0. \tag{3.11}$$

Es fácil demostrar que, cuando $\theta \in (\theta_0, \theta_1)$, la expresión entre paréntesis en (3.11) es negativa (véase la Sección 2.2). Por lo tanto, la desigualdad (3.11) es equivalente a

$$n > \left(\frac{\log \lambda_0}{\theta - \theta_0} + \frac{\log \lambda_1}{\theta_1 - \theta} \right) \Bigg/ \left(\frac{b(\theta_1) - b(\theta)}{\theta_1 - \theta} - \frac{b(\theta) - b(\theta_0)}{\theta - \theta_0} \right). \tag{3.12}$$

Denotemos por $H = H(\lambda_0, \lambda_1, \theta_0, \theta_1, \theta)$ el entero mínimo mayor que el lado derecho de (3.12). Se deduce que H es una cota superior para el número máximo de observaciones que puede requerir una prueba óptima $(\psi, \phi) \in \mathscr{M}_\theta$, cuando $\theta \in (\theta_0, \theta_1)$.

No existe un número máximo de pasos para una prueba $(\psi, \phi) \in \mathscr{M}_\theta$ si $\theta \notin (\theta_0, \theta_1)$.

Si $\theta > \theta_1$, entonces

$$\{s : \psi_n(s) < 1\} \subset \{s : s \leq \text{mín}\,\{B_n(1/\lambda_1, \theta, \theta_1),\ B_n(\lambda_0, \theta_0, \theta)\}\}. \tag{3.13}$$

Si $\theta < \theta_0$, entonces

$$\{s : \psi_n(s) < 1\} \subset \{s : s \geq \text{máx}\,\{B_n(1/\lambda_1, \theta, \theta_1),\ B_n(\lambda_0, \theta_0, \theta)\}\}. \tag{3.14}$$

Si $\theta = \theta_1$, entonces

$$\{s : \psi_n(s) < 1\} \subset \{s : s \leq B_n(\lambda_0, \theta_0, \theta_1)\}. \tag{3.15}$$

Finalmente, si $\theta = \theta_0$, entonces

$$\{s : \psi_n(s) < 1\} \subset \{s : s \geq B_n(1/\lambda_1, \theta_0, \theta_1)\}. \tag{3.16}$$

Para concluir esta parte, observemos que la regla de decisión óptima en (3.6) puede definirse en todos los casos como

$$\phi_n = I_{[B_n(\lambda_0/\lambda_1, \theta_0, \theta_1), \infty)}$$

para todo n natural (cf. la definición de B_n en (3.8) y la definición de ϕ_n en (3.6)).

3.2.2. Algoritmos computacionales para el diseño de pruebas óptimas

La base teórica está dada por las fórmulas (3.6) y (3.7).

Consideremos primero el caso en que $\theta \in (\theta_0, \theta_1)$.

Se deduce de (3.10) que el intervalo de continuación está acotado (o vacío) para

todo n natural. Por lo tanto, para evaluar todas las características de la prueba óptima, se necesita el conjunto de intervalos de continuación $[a_n, b_n]$, $n = 1, \ldots, H-1$, junto con los valores de las funciones $U^H_{\theta,n}$ en dichos intervalos.

Por supuesto, no es posible almacenar en la memoria de una computadora un continuo de valores de $U^H_{\theta,n}$, por lo que proponemos almacenar, en su lugar, sus valores sobre una malla adecuada en cada intervalo de continuación, utilizando interpolación para los demás puntos. De este modo, en lugar de trabajar con las funciones $U^H_{\theta,n}$ definidas por (3.2)–(3.4), trabajaremos con $\widetilde{U}^H_{\theta,n}(s)$, definidas para $n = H$ como $\widetilde{U}^H_{\theta,H}(s) \equiv U^H_{\theta,H}(s)$ y, recursivamente para $n = H, \ldots, 2$, como

$$\widetilde{U}^H_{\theta,n-1}(s) = \text{mín}\left\{\lambda_0 g^{n-1}_{\theta_0}(s),\ \lambda_1 g^{n-1}_{\theta_1}(s),\ g^{n-1}_{\theta}(s) + \mathcal{J}\widetilde{U}^H_{\theta,n}(s)\right\}, \quad s \in \mathbb{R}. \tag{3.17}$$

Sabemos por el apartado anterior que $H \leq H(\lambda_0, \lambda_1, \theta_0, \theta_1, \theta)$.

Así, una prueba óptima puede obtenerse "hacia atrás" desde $H = H(\lambda_0, \lambda_1, \theta_0, \theta_1, \theta)$ del siguiente modo:

1. Fijar $n = H$.

2. Definir a_{n-1} y b_{n-1}, respectivamente, como el mínimo y el máximo valor de $s \in [B_{n-1}(1/\lambda_1, \theta, \theta_1), B_{n-1}(\lambda_0, \theta_0, \theta)]$ para los cuales se cumple

$$g^{n-1}_{\theta}(s) + \mathcal{J}\widetilde{U}^H_{\theta,n}(s) \leq \text{mín}\left\{g^{n-1}_{\theta_0}(s),\ g^{n-1}_{\theta_1}(s)\right\}. \tag{3.18}$$

   Si no existe tal s, entonces fijar $H = n-1$, y si $H = 1$ entonces detenerse; en caso contrario, volver al Paso 1.

3. Para cada punto s en una malla de valores $\{s^{n-1}_i\}$ en $[a_{n-1}, b_{n-1}]$, evaluar y almacenar $v = v^{n-1}_i = g^{n-1}_{\theta}(s) + \mathcal{J}\widetilde{U}^H_{\theta,n}(s)$. Tener en cuenta, para uso posterior, que

$$\widetilde{U}^H_{\theta,n-1}(s) = \begin{cases} v^{n-1}_i + \dfrac{v^{n-1}_{i+1} - v^{n-1}_i}{s^{n-1}_{i+1} - s^{n-1}_i}(s - s^{n-1}_i), & \text{si } s \in [s^{n-1}_i, s^{n-1}_{i+1}] \\ \text{mín}\left\{g^{n-1}_{\theta_0}(s),\ g^{n-1}_{\theta_1}(s)\right\}, & \text{si } s \notin [a_{n-1}, b_{n-1}]. \end{cases}$$

4. Si $n = 2$ entonces detenerse; en caso contrario, fijar $n = n-1$ y volver al Paso 2.

Si $H = 1$ cuando el algoritmo se detiene, entonces la prueba óptima no es secuencial (se detiene después de la primera etapa). Usualmente, esto significa que λ_0 y/o λ_1 son demasiado pequeños para generar una prueba secuencial significativa. En este caso, una prueba óptima puede definirse como $\psi_1 \equiv 1$ y $\phi_1 = 1 - I_{[0, B_1(\lambda_0/\lambda_1, \theta_0, \theta_1)]}$.

Si $H > 1$, se obtiene una prueba óptima con intervalos de continuación $[a_n, b_n]$, $n = 1, 2, \ldots, H-1$, y un número máximo de observaciones igual a H. En cualquier caso, $\phi_n = 1 - I_{[0,B_n(\lambda_0/\lambda_1,\theta_0,\theta_1)]}$ puede aplicarse en cualquier etapa $n = 1, 2, \ldots, H$.

Si $\theta \notin (\theta_0, \theta_1)$, el algoritmo anterior puede utilizarse, con algunas modificaciones, para obtener la prueba óptima $\mathscr{M}_\theta^H$ dentro de la clase de pruebas *truncadas*. Las modificaciones se refieren a la forma en que se realiza la "inducción hacia atrás" en el Paso 2, ya que en cada uno de los casos considerados en la Subsection 3.2.1, cuando $\theta \notin (\theta_0, \theta_1)$, existe solo una cota para la región de continuación.

Por ejemplo, si $\theta > \theta_1$, se busca un valor de s que satisfaga (2.12) descendiendo desde la cota superior de la región de continuación (véase (3.13)); luego, una vez encontrado b_{n-1}, se continúa la búsqueda de a_{n-1} descendiendo.

De manera similar, cuando $\theta < \theta_0$, se busca un valor de s que satisfaga (2.12) ascendiendo desde la cota inferior dada en (3.14); luego, una vez encontrado a_{n-1}, se continúa la búsqueda de b_{n-1} ascendiendo.

En sentido estricto, en ninguno de los casos podemos estar seguros de obtener una prueba óptima de esta manera cuando $\theta \notin [\theta_0, \theta_1]$. Sí podríamos, si se supiera que la región de continuación de la prueba óptima debe ser un intervalo. Que sepamos, no existen resultados teóricos al respecto cuando $\theta \notin [\theta_0, \theta_1]$.

Los casos $\theta = \theta_0$ o $\theta = \theta_1$ también tienen fronteras unilaterales ((3.15) y (3.16)), pero se sabe que la región de continuación de la prueba óptima (SPRT) es un intervalo en estos casos; por lo tanto, utilizando el algoritmo anterior, podemos aproximarnos tanto como deseemos a la prueba óptima (haciendo suficientemente pequeña la malla), tanto en horizonte finito como (haciendo H suficientemente grande) en horizonte infinito.

3.2.3. Característica de operación, número promedio de observaciones y fórmulas relacionadas

En esta subsección obtenemos fórmulas para calcular las probabilidades de error, el número promedio de observaciones y algunas características relacionadas de pruebas secuenciales con una forma bastante general, es decir, aquellas basadas en sumas acumuladas de observaciones independientes, como las pruebas que obtuvimos anteriormente en el problema modificado de Kiefer–Weiss, pero también aplicables a pruebas como SPRT, 2-SPRT, etc.

Consideremos pruebas secuenciales basadas en sumas acumuladas S_1, $S_2, \ldots, S_n$ de observaciones independientes, suponiendo que las pruebas tienen la siguiente forma genérica.

Para un horizonte finito H, sean $a_1 < b_1$, $a_2 < b_2$, $\ldots$, $a_H = b_H$ algunos números reales.

Definimos una regla de paro ψ como

$$\psi_n = I_{\{S_n \notin (a_n, b_n)\}}, \tag{3.19}$$

y una regla de decisión ϕ como

$$\phi_n = I_{\{S_n \geq b_n\}}, \tag{3.20}$$

para $n = 1, 2, \ldots, H$.

En la versión con horizonte infinito, la prueba está definida por una sucesión $a_1 < b_1$, $a_2 < b_2$, ..., con ψ y ϕ que satisfacen (3.19) y (3.20), respectivamente, para todo n natural.

Lorden (1980) demostró que las pruebas óptimas en el problema modificado de Kiefer–Weiss para la familia de Koopman–Darmois uniparamétrica tienen esta forma cuando θ está estrictamente entre θ_0 y θ_1. Kiefer and Weiss (1957) proporcionaron condiciones más generales bajo las cuales las pruebas bayesianas son del tipo denominado SPRT generalizado; por lo tanto, están basadas en sumas acumuladas de variables aleatorias i.i.d. Las SPRTs truncadas también son del tipo (3.19)–(3.20) y pueden utilizarse para la evaluación aproximada de las características de las SPRTs al truncarlas con un H suficientemente grande. Las 2-SPRTs propuestas por Lorden (1976) también son del tipo (3.19)–(3.20) cuando se muestrea a partir de una familia exponencial.

A continuación, obtenemos fórmulas para evaluar las características numéricas de pruebas (ψ, ϕ) del tipo (3.19)–(3.20): el número promedio de observaciones $N(\theta; \psi)$ y la característica de operación $\mathrm{OC}_\theta(\psi, \phi)$. Haciendo $H \to \infty$ se obtiene un método para calcular también aproximaciones numéricas de pruebas con horizonte infinito, por ejemplo, las SPRTs.

Las constantes a_n, b_n, $n = 1, \ldots, H$, se consideran fijas, por lo que las omitiremos en la notación. Sea

$$\tau = \text{mín}\{n \geq 1 : S_n \notin (a_n, b_n)\}.$$

Definamos, para $n = 2, 3, \ldots, H-1$,

$$G_n(b, \theta) = \int_{a_1}^{b_1} f_\theta(y_1)\, dy_1 \int_{a_2}^{b_2} f_\theta(y_2 - y_1)\, dy_2 \cdots \int_{a_{n-1}}^{b_{n-1}} f_\theta(y_{n-1} - y_{n-2}) f_\theta(b - y_{n-1})\, dy_{n-1} \tag{3.21}$$

(suponiendo por definición que $y_0 = 0$), y definamos $G_1(b, \theta) = f_\theta(b)$, $b \in \mathbb{R}$, $\theta \in \Theta$.

Integrando ambos lados de (3.21) con respecto a b sobre el intervalo (a_n, b_n), obtenemos:

$$\begin{aligned}\int_{a_n}^{b_n} G_n(b,\theta)\,db &= \int_{a_1}^{b_1} f_\theta(y_1)\,dy_1 \int_{a_2}^{b_2} f_\theta(y_2-y_1)\,dy_2 \cdots \int_{a_n}^{b_n} f_\theta(b-y_{n-1})\,db \qquad (3.22)\\ &= P_\theta(S_1 \in (a_1,b_1), S_2 \in (a_2,b_2), \ldots, S_n \in (a_n,b_n))\\ &= P_\theta(\tau > n) \qquad (3.23)\end{aligned}$$

para todo $n = 1, \ldots, H-1$.

Se deduce de (3.22)–(3.23) que G_n puede utilizarse para calcular el número promedio de observaciones:

$$N(\theta;\psi) = \sum_{n=0}^{H-1} P_\theta(\tau > n) = 1 + \sum_{n=1}^{H-1} \int_{a_n}^{b_n} G_n(b,\theta)\,db. \qquad (3.24)$$

Desafortunadamente, no hay esperanza de obtener expresiones analíticas para G_n, al menos no en esta forma general. Por esta razón, proponemos calcularlas numéricamente, con ayuda de la siguiente fórmula recursiva.

Proposición 3.2.1 *Para todo $n = 1, 2, \ldots, H-2$ y para todo $y \in \mathbb{R}$ y $\theta \in \Theta$, se cumple que*

$$G_{n+1}(y,\theta) = \int_{a_n}^{b_n} G_n(s,\theta)\, f_\theta(y-s)\,ds.$$

La demostración de la Proposición 3.2.1 se deduce fácilmente de la definición (3.21) por inducción.

De este modo, comenzando desde $n = 1$ y utilizando una malla adecuada en cada intervalo (a_n, b_n), podemos evaluar aproximadamente G_n a partir de los valores en la malla de G_{n-1}, empleando una fórmula de cuadratura conveniente.

También existe un método análogo para calcular la característica de operación a partir de las funciones G_n.

Primero, integrando ambos lados de (3.21) con respecto a b sobre el intervalo $(-\infty, y)$, obtenemos:

Proposición 3.2.2 *Para todo $n = 1, 2, \ldots, H-1$ y para todo $y \in \mathbb{R}$ y $\theta \in \Theta$, se cumple que*

$$P_\theta(S_1 \in (a_1,b_1), \ldots, S_n \in (a_n,b_n),\ S_{n+1} < y) = \int_{a_n}^{b_n} G_n(s,\theta)\, F_\theta(y-s)\,ds,$$

donde F_θ es la función de distribución acumulada correspondiente a f_θ.

Aplicando la definición de la característica de operación en (1.1) a las expresiones (3.19)–(3.20), obtenemos

$$\mathrm{OC}_\theta(\psi,\phi) = P_\theta(S_1 < a_1) + \sum_{n=1}^{H-1} P_\theta(S_1 \in (a_1,b_1),\ldots,S_n \in (a_n,b_n),\ S_{n+1} < a_{n+1}),$$

lo cual justifica, en virtud de la Proposición 3.2.2, el siguiente resultado.

Corolario 3.2.3 *Para todo* $\theta \in \Theta$,

$$\mathrm{OC}_\theta(\psi,\phi) = F_\theta(a_1) + \sum_{n=1}^{H-1} \int_{a_n}^{b_n} G_n(s,\theta)\, F_\theta(a_{n+1}-s)\, ds \tag{3.25}$$

Para el cálculo numérico, pueden utilizarse los mismos valores en la malla de G_n que se emplearon anteriormente.

3.3. Resultados numéricos

En esta sección presentamos los resultados de las evaluaciones numéricas para el problema de Kiefer–Weiss con muestreo de una distribución normal con varianza conocida (sin pérdida de generalidad, asumimos que es igual a 1).

Para implementar todos los algoritmos descritos anteriormente, solo se necesitan unos pocos elementos específicos del modelo. En (2.1), se tiene $b(\theta) = \theta^2/2$ y $h(x) = \phi(x)$, la función de densidad de probabilidad de la normal estándar. En consecuencia, en todas las fórmulas de la subsección 3.2.1 se tiene

$$B_n(\lambda,\theta_0,\theta_1) = \frac{\log\lambda}{\theta_1-\theta_0} + n\frac{\theta_0+\theta_1}{2},$$

y la cota en el lado derecho de (2.11) es

$$\frac{2\left(\frac{\log\lambda_0}{\theta-\theta_0} + \frac{\log\lambda_1}{\theta_1-\theta}\right)}{\theta_1-\theta_0}.$$

El operador (3.4) es la parte más laboriosa del trabajo de programación para el diseño de la prueba óptima y, en el caso normal, tiene la forma

$$\mathcal{J}U(s) = \int_{\mathbb{R}} U(s+x)\,\phi(x)\,dx = \int_{\mathbb{R}} U(x)\,\phi(x-s)\,dx, \quad s \in \mathbb{R}. \tag{3.26}$$

Teniendo en cuenta que, en la implementación computacional, solo almacenamos los valores de U sobre una malla dentro del intervalo de continuación (véase (3.17) y el algoritmo en el subsección 3.2.2), únicamente la parte de (3.26) correspondiente al intervalo de continuación entra en consideración para la integración numérica, porque las otras dos partes pueden calcularse analíticamente. Se puede emplear

cualquier fórmula de cuadratura conveniente para la integración numérica en dicho intervalo.

Usando las especificaciones del caso normal como referencia, todos los algoritmos pueden adaptarse prácticamente a cualquier otra familia exponencial de distribuciones.

En nuestro ejemplo numérico, ponemos a prueba la hipótesis $\mathrm{H}_0\colon \theta = \theta_0 = -0.1$ frente a $\mathrm{H}_1\colon \theta = \theta_1 = 0.1$ sobre el valor medio θ de la distribución normal (este es el caso más laborioso de los estudiados en Lorden (1976)).

Se utiliza el método de la Sección 2.2 de Novikov et al. (2022) para reducir el problema de Kiefer–Weiss a su versión modificada. Usamos la Opción 1 del método, empleando la cota superior derivada de (2.11) para la distribución normal. Para resolver el problema modificado de Kiefer–Weiss para cualquier θ (como se requiere en (2.15) de Novikov et al. (2022)), utilizamos el algoritmo de la subsección 3.2.2 con una malla de tamaño 0.02 en cada intervalo $[a_n, b_n]$. Un tamaño de malla mayor permite realizar los cálculos más rápidamente, obteniendo resultados similares.

Para representar nuestros resultados de forma unificada, utilizamos tres niveles de probabilidades de error, 0.1, 0.05 y 0.01, para α y β, realizando los cálculos para cualquier combinación de α y β elegida entre estos valores. Para cumplir con las restricciones, para cada par de valores α y β buscamos los valores de λ_0 y λ_1 para los cuales la prueba correspondiente $\mathscr{M}_\theta$ proporcione probabilidades de error suficientemente cercanas a α y β; concretamente, exigimos que se cumplan las condiciones

$$\frac{|\alpha - \alpha(\mathscr{M}_\theta)|}{\alpha} < 0.002 \quad \text{y} \quad \frac{|\beta - \beta(\mathscr{M}_\theta)|}{\beta} < 0.002.$$

Esto puede parecer complicado, pero el siguiente hecho empírico lo simplifica bastante: el efecto de λ_0 sobre las probabilidades de error se concentra principalmente en α, dejando β prácticamente sin cambios, mientras que λ_1 afecta principalmente a β, sin modificar sustancialmente α. Además, el efecto es de sentido contrario: valores más grandes de λ_0 disminuyen α, y λ_1 produce el mismo efecto sobre β. Teniendo a disposición una implementación del algoritmo principal en forma de código computacional, se pueden encontrar fácilmente valores satisfactorios de λ_0 y λ_1 tras algunas pruebas por ensayo y error. Por supuesto, si se requiere una mayor precisión en las probabilidades de error, se puede adaptar fácilmente una rutina de optimización general para realizar el ajuste con mayor exactitud.

Presentamos los resultados del cálculo descrito en la Tabla 3.1. La primera parte (bloque superior) de la tabla contiene los parámetros y características de las pruebas óptimas en el problema de Kiefer–Weiss. Todos los valores numéricos de la tabla se obtuvieron mediante los cálculos descritos anteriormente, con el objetivo de alcanzar los valores fijos de α y β indicados en el encabezado. Solo se muestra una parte de

los resultados para el caso $\alpha \neq \beta$, concretamente aquellos con $\alpha > \beta$, ya que para el caso $\alpha < \beta$ los valores correspondientes pueden obtenerse intercambiando λ_0 y λ_1, $N(\theta_0; \mathscr{M}_\theta)$ y $N(\theta_1; \mathscr{M}_\theta)$, y cambiando el signo de θ, etc.

Los valores mostrados de λ_0 y λ_1 son aquellos que cumplen con las condiciones sobre α y β según lo explicado anteriormente. El valor de H se obtiene como resultado del algoritmo presentado en la Sección 3.2.2, en conjunto con el algoritmo general de la Sección 2.2 de Novikov et al. (2022). H representa el número máximo de observaciones que puede requerir la prueba óptima, y θ es el valor que se utiliza en el problema modificado al que se reduce el problema de Kiefer–Weiss.

Estos parámetros solo son necesarios para garantizar la reproducibilidad de las evaluaciones de los planes de muestreo óptimos, una vez que se dispone de una rutina de optimización para el problema modificado de Kiefer–Weiss. También se presentan en la tabla las características de la prueba óptima: $N(\theta; \mathscr{M}_\theta)$, $N(\theta_0; \mathscr{M}_\theta)$, $N(\theta_1; \mathscr{M}_\theta)$, y el percentil 99 del número de observaciones que toma la prueba óptima $\mathscr{M}_\theta$ cuando el valor verdadero del parámetro es θ. Este valor se indica en la tabla como $Q_{99}(\theta; \mathscr{M}_\theta)$.

El valor reportado de $\Delta(\mathscr{M}_\theta)$ no forma parte del procedimiento de construcción de la prueba óptima (a diferencia de θ y H), y se incluye únicamente para mostrar que tan cercano está el número promedio de observaciones $N(\theta; \mathscr{M}_\theta)$ a su valor máximo $\sup_\vartheta N(\vartheta; \mathscr{M}_\theta)$ (el valor de Δ debe ser cero para una prueba exactamente óptima).

En la segunda parte de la Tabla 3.1 (bloque central) se presentan las características de la SPRT cuyas probabilidades de error coinciden con los valores correspondientes de α y β. Los extremos (A, B) del intervalo de continuación se obtienen utilizando la función `fminsearch` del paquete `neldemead` de R para optimización sin derivadas (Bihorel, 2022). Para cada par A, B, usamos la fórmula (3.24) para evaluar aproximadamente el número promedio de observaciones, tomando un valor de H lo suficientemente grande como para que la probabilidad de continuar después de H pasos sea menor que 10^{-4}. Este valor se muestra en la Tabla 3.1 como $H_{\text{trunc.}}$ Dado que se sabe que la SPRT tiende a requerir un número elevado de observaciones cuando θ está entre θ_0 y θ_1, utilizamos el valor de θ correspondiente al problema de Kiefer–Weiss al determinar $H_{\text{trunc.}}$ Este número luego se utiliza para evaluar las probabilidades de error mediante la fórmula (3.25) y para calcular el número promedio de observaciones bajo H_0 y H_1. Por la misma razón, se usa el mismo valor de θ para calcular el número promedio de observaciones $N(\theta; W)$ y el percentil 99 $Q_{99}(\theta; W)$ de la SPRT.

Finalmente, FSS (la fila anterior al bloque inferior) representa el valor mínimo del tamaño de muestra requerido por la prueba óptima con tamaño fijo, con

probabilidades de error α y β. Para valores dados de α y β, FSS se calcula como

$$n^* + \frac{\beta(n^*) - \beta}{\beta(n^*) - \beta(n^* + 1)},$$

donde $n^* = n^*(\alpha, \beta)$ es el entero máximo n tal que la probabilidad de error de tipo II, $\beta(n)$, de la prueba más poderosa con nivel α (prueba de Neyman–Pearson) es mayor o igual a β.

En el último bloque de la tabla calculamos los valores de eficiencia relativa para facilitar la comparación entre las tres pruebas. Tomamos como referencia el valor FSS en todas las comparaciones y definimos

$$R(\mathscr{M}_\theta) = \frac{FSS}{N(\theta; \mathscr{M}_\theta)}, \quad R_i(\mathscr{M}_\theta) = \frac{FSS}{N(\theta_i; \mathscr{M}_\theta)},\ i = 0, 1, \quad \text{y} \quad RQ(\mathscr{M}_\theta) = \frac{FSS}{Q_{99}(\theta; \mathscr{M}_\theta)}$$

para la prueba óptima de Kiefer–Weiss, y

$$R(W) = \frac{FSS}{N(\theta; W)}, \quad R_i(W) = \frac{FSS}{N(\theta_i; W)},\ i = 0, 1, \quad \text{y} \quad RQ(W) = \frac{FSS}{Q_{99}(\theta; W)}$$

para la SPRT.

Por ejemplo, si $R_0(\mathscr{M}_\theta) = 2$, esto significa que la prueba con tamaño de muestra fijo requiere, en promedio bajo H_0, el doble de observaciones que la prueba óptima de Kiefer–Weiss.

Se ha llevado a cabo otra serie de evaluaciones en un rango más amplio de valores de α y β (aproximadamente entre 0.1 y 0.0005), con el fin de analizar el efecto de α y β sobre todos los elementos y características de la prueba óptima de Kiefer–Weiss. Los datos corresponden a la misma hipótesis sobre el valor medio de la distribución normal: $\theta_0 = -0.1$ frente a $\theta_1 = 0.1$, asumiendo varianza unitaria.

Presentamos los resultados en forma gráfica en las Figuras 3.1 a 3.7. Todos los gráficos se basan en una malla de 8×8 valores de λ_0 y λ_1, que genera una cuadrícula casi uniforme de valores de α y β en escala logarítmica.

Para cada par fijo de λ_0 y λ_1 considerados en los cálculos, se obtienen valores de $\Delta = \Delta(\mathscr{M}_\theta)$ muy cercanos a 0 (hasta 10^{-4}, dado el tamaño de malla de 0.02), lo cual se interpreta como una solución numérica al problema de Kiefer–Weiss mediante $\mathscr{M}_\theta$, con el nivel de precisión dado.

Para una mejor visualización, se ajusta un modelo de regresión polinómica local (LOESS)[1] para representar cómo dependen las características de las pruebas óptimas de los valores de α y β. En todos los gráficos se utilizan los logaritmos decimales de

[1] https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/loess

α y β.

La Figura 3.1 representa un gráfico de contorno correspondiente al número promedio de observaciones $N(\theta; \mathscr{M}_\theta)$ como función de α y β. Los valores calculados de $N(\theta; \mathscr{M}_\theta)$ varían, de manera muy simétrica respecto a $\alpha = \beta$, desde aproximadamente 150 hasta casi 900 observaciones. Como era de esperarse, el número máximo corresponde a los valores mínimos de $\alpha = \beta$ utilizados.

La Figura 3.2 muestra el comportamiento de la eficiencia relativa $R(\theta; \mathscr{M}_\theta) = FSS/N(\theta; \mathscr{M}_\theta)$ como función de α y β. Tal como en otros casos evaluados anteriormente en los secciones previos, el valor mínimo tiende a alcanzarse cuando α y β son pequeños e iguales (el mínimo teórico absoluto de esta eficiencia es 1, pero observamos que para los valores de α y β que utilizamos, este mínimo es aproximadamente 1.28).

Más importante desde el punto de vista práctico, los datos de eficiencia se presentan en las Figuras 3.3 y 3.4. Estas figuras se refieren a $R_0(\mathscr{M}_\theta)$ y $R_1(\mathscr{M}_\theta)$ y muestran cuánto se gana con los planes secuenciales en comparación con el tamaño fijo de muestra, bajo H_0 y H_1, respectivamente. Tal como en otros casos estudiados en los secciones anteriores, la eficiencia aumenta cuando α y β se vuelven muy asimétricos. $R_0(\mathscr{M}_\theta)$ aumenta cuando α es pequeña en comparación con β, y $R_1(\mathscr{M}_\theta)$ aumenta, de forma perfectamente simétrica, cuando β es pequeña en comparación con α (hasta aproximadamente 4 en ambos casos, para el rango de valores de α y β utilizados). La eficiencia en el peor caso en el escenario asimétrico es aproximadamente 1.6, lo cual sigue siendo mejor que cuando las hipótesis están completamente mal especificadas (Figura 3.2). En el caso en que $\alpha \approx \beta$, ambos coeficientes de eficiencia aumentan lentamente desde aproximadamente 1.9 hasta 2.4 cuando disminuyen las probabilidades de error.

La Figura 3.5 se entiende mejor en conjunto con otros datos sobre la ubicación de los cuantiles de la distribución del tamaño muestral con respecto a FSS, como por ejemplo con $RQ(\mathscr{M}_\theta)$ y $RQ(W)$ en la Tabla 3.1.

Así, se observa en la Figura 3.5 que FSS se sitúa en el nivel de aproximadamente 60 % a 80 % del percentil 99 de la distribución del tamaño muestral, con una clara tendencia a aumentar cuando α y β son pequeños. Niveles más altos de esta razón indican que el tamaño muestral raramente excede el nivel de FSS.

Como se ve en la Tabla 3.1, para niveles usuales de α y β, esta razón es aproximadamente 60 % para las pruebas óptimas de Kiefer-Weiss, mientras que para SPRT es de alrededor del 30 %, lo que significa que la SPRT emplea, bajo el mismo valor de θ y con la misma probabilidad de 0.01, un tamaño muestral aproximadamente tres veces mayor que el de FSS. De hecho, este tipo de comportamiento de la SPRT fue la razón por la que se planteó el problema que estamos abordando ahora (véase Kiefer and Weiss, 1957).

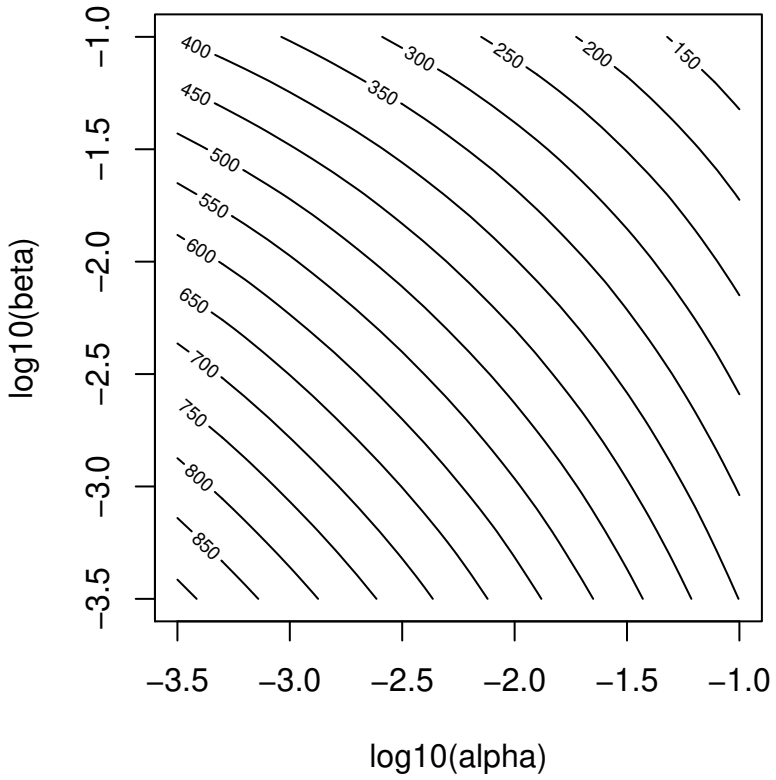


Figura 3.1: Prueba óptima de Kiefer-Weiss: número promedio de observaciones bajo θ

Es notable que el mismo patrón en cuanto al comportamiento de los percentiles se haya observado en todos los modelos que investigamos anteriormente (véase el capítulo 2).

La Figura 3.6 se presenta simplemente para observar cómo depende el valor de θ, que hace óptima la solución modificada de Kiefer-Weiss, de los valores de α y β. Como era de esperarse, $\theta = 0$ cuando $\alpha = \beta$, y cambia de signo cuando se intercambian los valores de α y β. El valor máximo de $|\theta|$ es aproximadamente 0.05, alcanzando este máximo para valores muy asimétricos de α y β (recordemos que en nuestro ejemplo $\theta_1 = -\theta_0 = 0.1$).

Finalmente, la Figura 3.7 muestra la ubicación de la FSS con respecto al número máximo de observaciones H que requiere la prueba óptima de Kiefer-Weiss. Para valores usuales de α y β, la FSS se encuentra aproximadamente entre el 40 % y el 60 % del valor máximo H, con una clara tendencia creciente; es decir, H se aproxima a la FSS a medida que α y β disminuyen.

	Simétrica α, β			Asimétrica α, β		
$(\alpha,\ \beta)$	(0.10, 0.10)	(0.05, 0.05)	(0.01, 0.01)	(0.10, 0.05)	(0.10, 0.01)	(0.05, 0.01)
λ_0	530.3	1201.0	6923.5	621.0	807.5	1517.0
λ_1	530.3	1201.0	6923.5	1042.0	4999.0	5682.0
θ	0	0	0	-0.016	-0.035	-0.021
H	749	913	1263	846	1085	1110
$N(\theta;\mathscr{M}_\theta)$	113.70	192.43	402.23	150.06	231.84	286.30
$N(\theta_0;\mathscr{M}_\theta)$	90.99	140.25	250.48	123.21	197.99	220.46
$N(\theta_1;\mathscr{M}_\theta)$	90.99	140.25	250.48	105.34	124.20	163.33
$Q_{99}(\theta;\mathscr{M}_\theta)$	277	416	738	346	497	572
$\Delta(\mathscr{M}_\theta)$	0.0000	0.0000	0.0000	0.0036	0.0019	1.4E-5
log(A)	-2.080	-2.827	-4.479	-2.774	-4.384	-4.436
log(B)	2.080	2.827	4.479	2.133	2.176	2.870
$H_{\text{trunc.}}$	924	1660	4045	1262	2178	2706
$N(\theta;W)$	120.85	216.83	528.14	162.75	258.17	340.10
$N(\theta_0;W)$	88.10	132.70	225.44	119.01	191.31	209.04
$N(\theta_1;W)$	88.10	132.70	225.44	99.87	111.50	145.80
$Q_{99}(\theta;W)$	474	851	2074	646	1101	1380
FSS	164.24	270.55	541.15	214.10	325.08	394.39
$R(\mathscr{M}_\theta)$	1.44	1.41	1.35	1.43	1.40	1.38
$R_0(\mathscr{M}_\theta)$	1.80	1.93	2.16	1.74	1.64	1.79
$R_1(\mathscr{M}_\theta)$	1.80	1.93	2.16	2.03	2.62	2.42
$RQ(\mathscr{M}_\theta)$	0.59	0.65	0.73	0.62	0.65	0.69
$R(W)$	1.36	1.25	1.02	1.31	1.26	1.16
$R_0(W)$	1.86	2.04	2.40	1.80	1.70	1.89
$R_1(W)$	1.86	2.04	2.40	2.14	2.92	2.70
$RQ(W)$	0.35	0.32	0.26	0.33	0.30	0.29

Tabla 3.1: Comparación entre las pruebas óptimas de Kiefer-Weiss, SPRT y FSST para distribución normal $\theta_0 = -0.1$, $\theta_1 = 0.1$

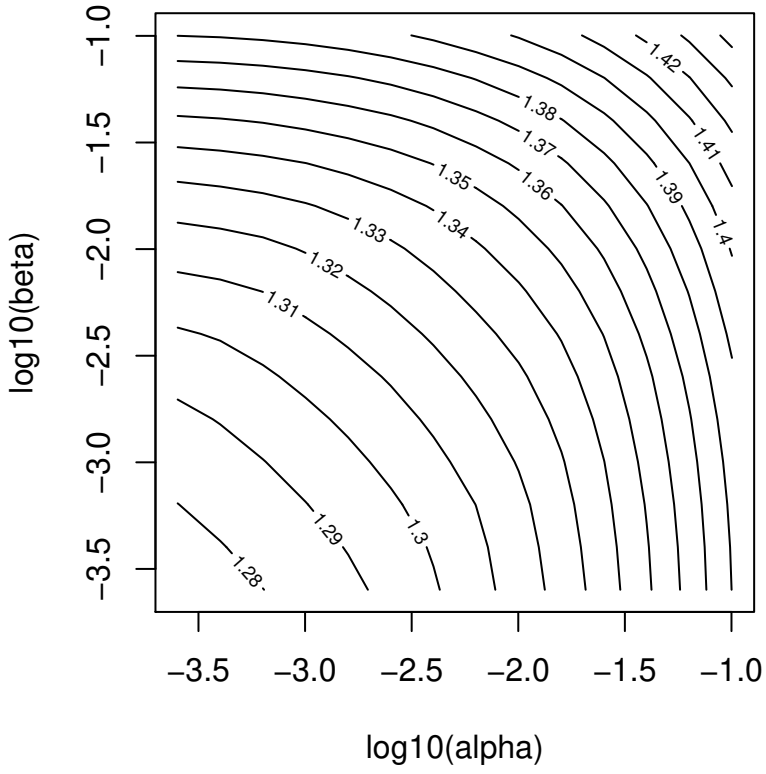


Figura 3.2: Prueba óptima de Kiefer-Weiss: FSS frente al número promedio de observaciones bajo θ

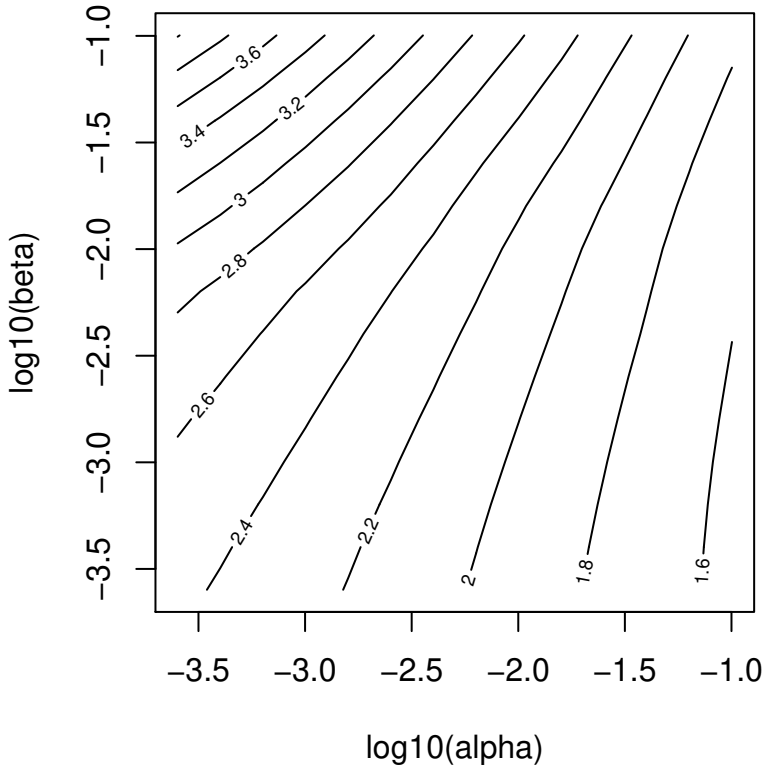


Figura 3.3: Prueba óptima de Kiefer-Weiss: FSS frente al número promedio de observaciones bajo H_0

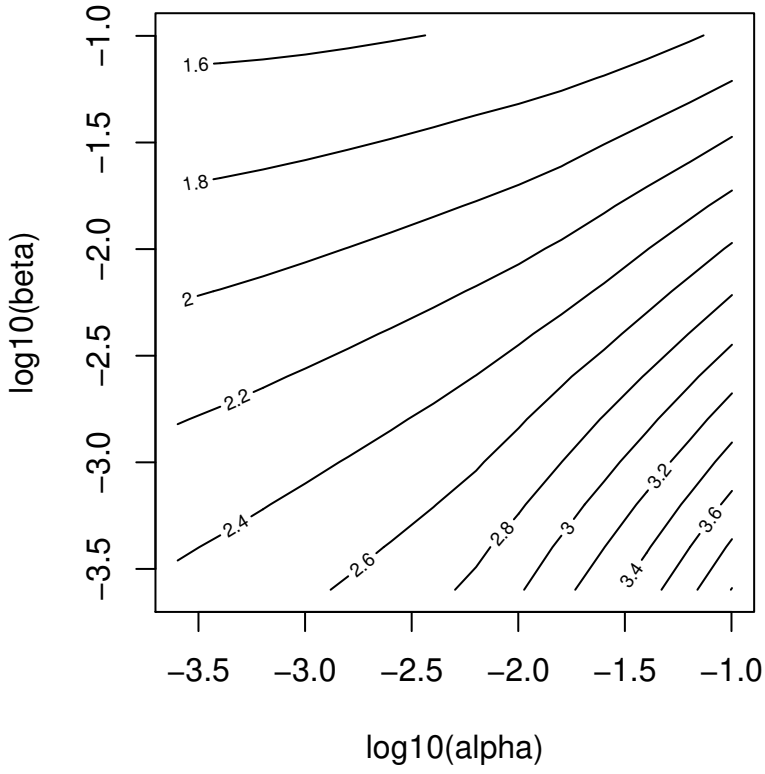


Figura 3.4: Prueba óptima de Kiefer-Weiss: FSS frente al número promedio de observaciones bajo H_1

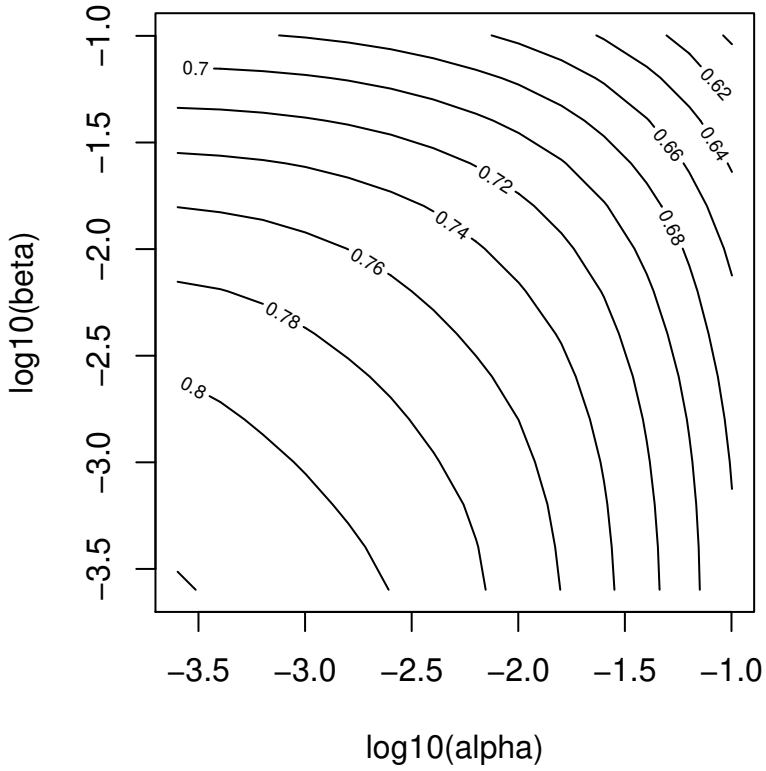


Figura 3.5: Prueba óptima de Kiefer-Weiss: FSS frente al percentil 99 del número de observaciones bajo θ

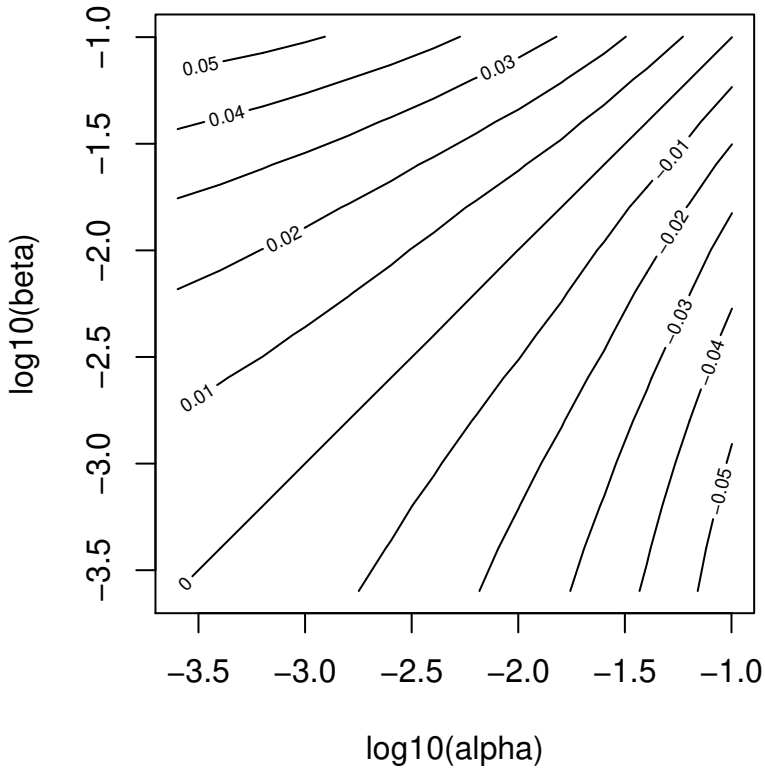


Figura 3.6: Prueba óptima de Kiefer-Weiss: θ como función de α y β

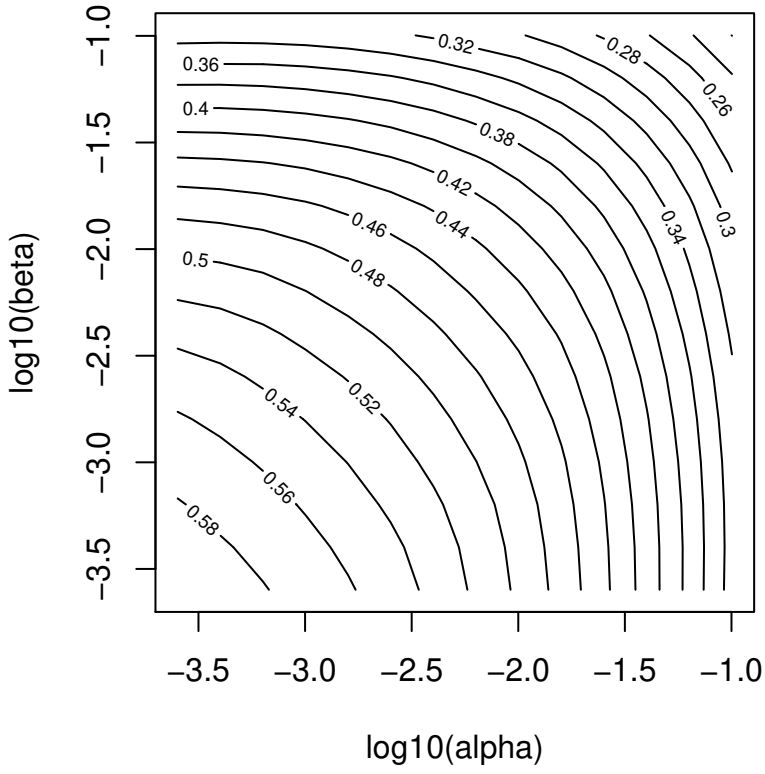


Figura 3.7: Prueba óptima de Kiefer-Weiss: FSS frente al número máximo de observaciones requerido

Capítulo 4

Pruebas secuenciales en grupos

4.1. Planes de muestreo óptimos

4.1.1. Pruebas secuenciales en grupos

En esta parte, formulamos el problema de diseño óptimo de pruebas secuenciales por grupos y describimos la estructura de las pruebas secuenciales por grupos óptimas.

Suponemos en este trabajo que los datos para el procedimiento de pruebas secuenciales por grupos llegan para su análisis en grupos de tamaños fijos, es decir, en cada etapa $n = 1, 2, \ldots, N$ se observa una cantidad predefinida m_n de datos, donde N es un número natural dado. Denotamos $m(n) = (m_1, \ldots, m_n)$, y sea $P_\theta^{m(n)}$ la distribución de las observaciones en los primeros n grupos, donde $\theta \in \Theta$ es un parámetro desconocido, $n = 1, 2, \ldots, N$. Nos ocupamos de contrastar una hipótesis simple $H_0 : \theta = \theta_0$ frente a una alternativa simple $H_1 : \theta = \theta_1$. Típicamente, los datos en el grupo n corresponden a una muestra aleatoria de tamaño m_n de una distribución paramétrica, pero también pueden considerarse como una estadística (suficiente) basada en dicha muestra.

Sean $y_1^{(m_1)}, y_2^{(m_2)}, \ldots, y_N^{(m_N)}$ los datos observados en los grupos $1, 2, \ldots, n$ (los correspondientes a $P_\theta^{m(n)}$), para $n = 1, \ldots, N$. Definimos una prueba secuencial por grupos como un par (ψ, ϕ) compuesto por una *regla de detención* ψ y una *regla de decisión* ϕ, donde $\psi = (\psi_1, \psi_2, \ldots, \psi_N)$ y $\phi = (\phi_1, \phi_2, \ldots, \phi_N)$, y las funciones

$$\psi_n = \psi_n(y_1^{(m_1)}, y_2^{(m_2)}, \ldots, y_n^{(m_n)}), \quad \phi_n = \phi_n(y_1^{(m_1)}, y_2^{(m_2)}, \ldots, y_n^{(m_n)})$$

se suponen medibles y con valores en $\{0, 1\}$. El valor $\psi_n = 1$ indica que se detiene en la etapa n, y el valor $\phi_n = 1$ (o 0) indica que se rechaza (o se acepta, respectivamente) la hipótesis H_0. Todos los algoritmos que desarrollamos en este trabajo pueden

extenderse fácilmente al caso más general de reglas aleatorizadas (cuando ψ_n y ϕ_n son distribuciones de probabilidad sobre $[0,1]$) (ver, por ejemplo, Ferguson, 1967).

Solo consideramos en este trabajo las pruebas secuenciales por grupos (ψ, ϕ) tales que $(1-\psi_1)\dots(1-\psi_N) \equiv 0$, es decir, aquellas que se detienen en o antes de la etapa N. Denotamos por $\mathscr{S}^N$ al conjunto de todas esas pruebas.

Ejemplo: una prueba en dos etapas para la media normal. *Supongamos que se puede tomar una muestra de n_1 observaciones en la primera etapa, y una de n_2 en la segunda. Se asume que todas las observaciones son independientes dentro de cada grupo y entre los grupos, y que tienen distribución normal $N(\theta, 1)$. Sea $\theta = \theta_0$ bajo la hipótesis H_0 y $\theta = \theta_1$ bajo H_1.*

Definamos un procedimiento de prueba en dos etapas que se detiene y acepta H_0 cuando la suma de la primera muestra S_1 es menor o igual que a_1, y se detiene y rechaza H_0 si S_1 es mayor o igual que b_1, donde $a_1 < b_1$ son constantes dadas. En el caso en que $a_1 < S_1 < b_1$, el procedimiento no se detiene y espera la segunda muestra.

En la segunda etapa, la prueba puede utilizar, además, la suma de la segunda muestra S_2. La prueba acepta H_0 cuando $S_1+S_2 < a_2$ y rechaza H_0 en caso contrario, deteniéndose en ambos casos.

Usando la notación $I\{A\}$ para la función indicadora de un evento A, tenemos $\psi_1 = I\{S_1 \le a_1\}+I\{S_1 \ge b_1\}$ y $\psi_2 \equiv 1$, por lo tanto, la regla de paro es $\psi = (\psi_1, \psi_2)$. La regla de decisión es $\phi = (\phi_1, \phi_2)$, donde $\phi_1 = I\{S_1 \ge b_1\}$ y $\phi_2 = I\{S_1+S_2 \ge a_2\}$.

Para cualquier regla de paro ψ, definimos

$$s_n^\psi = s_n^\psi(y_1^{(m_1)}, y_2^{(m_2)}, \dots, y_n^{(m_n)}) = (1-\psi_1)(1-\psi_2)\dots(1-\psi_{n-1})\psi_n.$$

Obviamente, $s_n^\psi = 1$ si y solo si la prueba se detiene en la etapa n.

Para una prueba secuencial $(\psi, \phi) \in \mathscr{S}^N$, las *probabilidades de error de tipo I y tipo II* se definen, respectivamente, como

$$\alpha(\psi, \phi) = \sum_{n=1}^{N} E_{\theta_0} s_n^\psi \phi_n \quad \text{y} \quad \beta(\psi, \phi) = \sum_{n=1}^{N} E_{\theta_1} s_n^\psi (1-\phi_n),$$

Es común expresar las probabilidades de error en términos de la característica de operación

$$\mathrm{OC}_\theta(\psi, \phi) = \sum_{n=1}^{N} E_\theta s_n^\psi (1-\phi_n).$$

De este modo, $\alpha(\psi, \phi) = 1 - \mathrm{OC}_{\theta_0}(\psi, \phi)$ y $\beta(\psi, \phi) = \mathrm{OC}_{\theta_1}(\psi, \phi)$.

La forma tradicional de tener en cuenta la cantidad de datos utilizados para la

inferencia en el análisis secuencial es calcular el número promedio de observaciones

$$\sum_{n=1}^{\infty} (m_1 + m_2 + \cdots + m_n)\, E_\theta s_n^\psi.$$

En las últimas décadas, ha ganado atención en la literatura un enfoque más general que considera el costo de las observaciones de una manera distinta a simplemente proporcional al número total de observaciones, es decir, suponiendo que el costo de obtener un grupo de m observaciones es $c(m)$ (por ejemplo, $c(m) = a + bm$, con $a > 0$) (ver, por ejemplo, Schmitz, 1993). En este trabajo, definimos el costo de las observaciones como $c_n(m)$ si se toman m observaciones en el grupo n. Esto puede ser útil al comparar diferentes planes de muestreo.

Definamos entonces el costo muestral esperado correspondiente a una regla de paro ψ, cuando θ es el valor verdadero del parámetro, como

$$C(\theta; \psi) = \sum_{n=1}^{N} \left(\sum_{i=1}^{n} c_i(m_i) \right) E_\theta s_n^\psi.$$

Definamos ahora el costo ponderado que queremos utilizar como función objetivo al buscar los planes óptimos. Sea $\vartheta = (\vartheta_1, \ldots, \vartheta_K)$ un conjunto de puntos del parámetro, y $\gamma = (\gamma_1, \ldots, \gamma_K)$ un conjunto de pesos positivos tal que $\sum_{i=1}^{K} \gamma_i = 1$, con $K \geq 1$.

Entonces, el costo esperado ponderado de muestreo $K(\vartheta, \gamma; \psi)$ se define como

$$K(\vartheta, \gamma; \psi) = \sum_{i=1}^{K} \gamma_i C(\vartheta_i; \psi),$$

El problema de optimización que abordamos en este trabajo consiste en minimizar $K(\vartheta, \gamma; \psi)$, dados ϑ y γ, dentro de la clase de pruebas $(\psi, \phi) \in \mathscr{S}^N$, sujeto a las restricciones

$$\alpha(\psi, \phi) \leq \alpha \text{ y } \beta(\psi, \phi) \leq \beta. \tag{4.1}$$

donde α, $\beta \in [0, 1]$ son números fijos.

La solución de este problema con restricciones puede obtenerse aplicando la idea del método de los multiplicadores de Lagrange, reduciendo así el problema de optimización con restricciones a uno sin restricciones.

Definamos la función de Lagrange

$$L(\psi, \phi) = L(\psi, \phi; \lambda_0, \lambda_1) = K(\vartheta, \gamma; \psi) + \lambda_0 \alpha(\psi, \phi) + \lambda_1 \beta(\psi, \phi), \tag{4.2}$$

donde λ_0, λ_1 son constantes no negativas (llamadas multiplicadores). Entonces, es

fácil ver que cualquier prueba (ψ, ϕ) que minimice (4.2) para algunos valores de λ_0, λ_1, también minimiza $K(\vartheta, \gamma; \psi)$ bajo las restricciones (4.1), siempre que se cumpla $\alpha(\psi, \phi) = \alpha$ y $\beta(\psi, \phi) = \beta$ (ver, por ejemplo, Novikov, 2009a).

De este modo, nuestro problema se reduce a la minimización de la función de Lagrange (4.2) en la clase $\mathscr{S}^N$ de pruebas secuenciales por grupos, dados ϑ y γ. Muchos criterios de optimización en la prueba de hipótesis secuencial por grupos pueden tratarse como casos particulares de la minimización de (4.2) (ver, por ejemplo, Jennison, 1987; Eales and Jennison, 1992).

En términos generales, muchos problemas secuenciales se reducen, o están relacionados, con alguna variante del problema de *paro óptimo* (Chow et al., 1971). Un algoritmo particular de Novikov (2009a) ha demostrado ser simple y útil para problemas de pruebas secuenciales, especialmente al trabajar con familias exponenciales de distribuciones (véase los capítulos 2 y 3). Adaptado al contexto secuencial por grupos, su descripción breve es la siguiente.

Sea $f_\theta^{m(n)} = f_\theta^{m(n)}(y_1^{(m_1)}, y_2^{(m_2)}, \ldots, y_n^{(m_n)})$ la densidad (derivada de Radon–Nikodym) de la distribución $P_\theta^{m(n)}$ con respecto a una medida producto $\mu^{m(n)} = \mu_{m_1} \times \cdots \times \mu_{m_n}$.

Partiendo de

$$V_N^{m(N)} = \text{mín}\{\lambda_0 f_{\theta_0}^{m(N)}, \lambda_1 f_{\theta_1}^{m(N)}\}, \tag{4.3}$$

definimos recursivamente para $n = N, \ldots, 2$:

$$V_{n-1}^{m(N)} = \text{mín}\{\lambda_0 f_{\theta_0}^{m(n-1)}, \lambda_1 f_{\theta_1}^{m(n-1)}, c_n(m_n) \sum_{i=1}^{K} \gamma_i f_{\vartheta_i}^{m(n-1)} + \mathcal{I}_{m_n} V_n^{m(N)}\}, \tag{4.4}$$

donde, para cualquier función medible no negativa $V(y_1^{(m_1)}, \ldots, y_n^{(m_n)})$, el operador $\mathcal{I}_{m_n} V$ se define como

$$\mathcal{I}_{m_n} = (\mathcal{I}_{m_n} V)(y_1^{(m_1)}, \ldots, y_{n-1}^{(m_{n-1})}) = \int V(y_1^{(m_1)}, \ldots, y_{n-1}^{(m_{n-1})}, y_n^{(m_n)}) d\mu_{m_n}(y_n^{(m_n)}). \tag{4.5}$$

Entonces, el mínimo de la función de Lagrange (4.2) igual a

$$c_1(m_1) + \mathcal{I}_{m_1} V_1^{m(N)} \tag{4.6}$$

y se alcanza en el par (ψ, ϕ) definido de la siguiente manera (donde $I\{A\}$ denota la función indicadora del evento A):

$$\psi_n = I\left\{ \text{mín}\{\lambda_0 f_{\theta_0}^{m(n)}, \lambda_1 f_{\theta_1}^{m(n)}\} \leq c_{n+1}(m_{n+1}) \sum_{i=1}^{K} \gamma_i f_{\vartheta_i}^{m(n)} + \mathcal{I}_{m_{n+1}} V_{n+1}^{m(N)} \right\} \tag{4.7}$$

para $n = 1, 2, \ldots, N-1$, y $\psi_N \equiv 1$, y

$$\phi_n = I\left\{\lambda_0 f_{\theta_0}^{m(n)} \leq \lambda_1 f_{\theta_1}^{m(n)}\right\} \tag{4.8}$$

para $n = 1, 2, \ldots, N$.

La demostración se lleva a cabo fácilmente siguiendo las líneas de Novikov (2009a).

4.1.2. Pruebas secuenciales óptimas por grupos para familias exponenciales

En esta sección, caracterizamos las pruebas secuenciales óptimas por grupos descritas en la sección anterior, en el caso en que los grupos están formados por muestras independientes de una distribución perteneciente a una familia exponencial de un solo parámetro. Más específicamente, supongamos que son muestras de una distribución cuya densidad (derivada de Radon–Nikodym con respecto a la medida de Lebesgue sobre $\mathbb{R}$) tiene la siguiente forma:

$$f_\theta(x) = \exp(\theta x - b(\theta))h(x),\ x \in \mathbb{R},\ \theta \in \Theta, \tag{4.9}$$

donde $h(x)$ es una función medible no negativa sobre $\mathbb{R}$, y Θ es el espacio natural del parámetro:

$$\Theta = \{\theta \in \mathbb{R} : \int_{\mathbb{R}} \exp(\theta x)h(x)dx < \infty\}.$$

Suponemos que la distribución es regular, lo cual significa que Θ es un subconjunto abierto de $\mathbb{R}$. También asumimos que $h(x)$ pertenece a la familia (y corresponde a $\theta = 0$).

Se pueden considerar familias exponenciales más generales mediante reparametrizaciones (definiendo θ como una función de un nuevo parámetro). Todos nuestros resultados a continuación también son aplicables a este caso, bajo supuestos naturales sobre la reparametrización.

Para los datos de los primeros n grupos, $y_1^{(m_1)}, y_2^{(m_2)}, \ldots, y_n^{(m_n)}$, la función de densidad conjunta es

$$f_\theta^{m(n)}(y_1^{(m_1)}, \ldots, y_n^{(m_n)}) = \exp\left(\theta \sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij} - b(\theta)\sum_{i=1}^{n} m_i\right)\prod_{i=1}^{n}\prod_{j=1}^{m_i} h(y_{ij}), \tag{4.10}$$

Denotemos por $h*h$ la convolución de h consigo misma, es decir,

$$h*h(x) = \int_{\mathbb{R}} h(y-x)\,h(y)\,dy,$$

y recursivamente $h^{*(n+1)} = h * (h^{*n})$, $n = 1, 2, \ldots$, comenzando con $h^{*1} \equiv h$.

Sea el operador $\mathcal{J}_m$ definido como

$$\mathcal{J}_m U(s) = \int_{\mathbb{R}} U(s+x)\, h^{*m}(x)\, dx, \quad s \in \mathbb{R}, \tag{4.11}$$

para cualquier función medible no negativa $U : \mathbb{R} \to \mathbb{R}$.

Denotemos

$$g_\theta^{m(n)}(s) = \exp\left(\theta s - b(\theta) \sum_{i=1}^{n} m_i\right), \quad s \in \mathbb{R},$$

y, comenzando con

$$U_N^{m(N)}(s) = \text{mín}\{\lambda_0 g_{\theta_0}^{m(N)}(s), \lambda_1 g_{\theta_1}^{m(N)}(s)\}, \quad s \in \mathbb{R}, \tag{4.12}$$

definimos recursivamente para $n = N, \ldots, 1$ las funciones

$$U_{n-1}^{m(N)}(s) = \text{mín}\left\{\lambda_0 g_{\theta_0}^{m(n-1)}(s), \lambda_1 g_{\theta_1}^{m(n-1)}(s), c_n(m_n) \sum_{i=1}^{K} \gamma_i g_{\vartheta_i}^{m(n-1)}(s) + \mathcal{J}_{m_n} U_n^{m(N)}(s)\right\}. \tag{4.13}$$

Entonces, las fórmulas de inducción hacia atrás (4.3)–(4.4) son equivalentes a (4.12)–(4.13), como se muestra en la siguiente proposición.

Proposición 4.1.1

$$V_n^{m(N)}(y_1^{(m_1)}, \ldots, y_n^{(m_n)}) = U_n^{m(N)}\left(\sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij}\right) \prod_{i=1}^{n}\prod_{j=1}^{m_i} h(y_{ij}), \tag{4.14}$$

para $n = 1, \ldots, N$.

Demostración: Por inducción sobre $n = N, N-1, \ldots, 1$.

A partir de (4.3), se tiene:

$$\begin{aligned} V_N^{m(N)}(y_1^{(m_1)}, \ldots, y_N^{(m_N)}) &= \text{mín}\{\lambda_0 f_{\theta_0}^{m(N)}(y_1^{(m_1)}, \ldots, y_N^{(m_N)}), \lambda_1 f_{\theta_1}^{m(N)}(y_1^{(m_1)}, \ldots, y_N^{(m_N)})\} \\ &= \text{mín}\left\{\lambda_0 g_{\theta_0}^{m(N)}\left(\sum_{i=1}^{N}\sum_{j=1}^{m_i} y_{ij}\right), \lambda_1 g_{\theta_1}^{m(N)}\left(\sum_{i=1}^{N}\sum_{j=1}^{m_i} y_{ij}\right)\right\} \prod_{i=1}^{N}\prod_{j=1}^{m_i} h(y_{ij}) \\ &= U_N^{m(N)}\left(\sum_{i=1}^{N}\sum_{j=1}^{m_i} y_{ij}\right) \prod_{i=1}^{N}\prod_{j=1}^{m_i} h(y_{ij}), \end{aligned}$$

por lo tanto, (4.14) se cumple para $n = N$.

Supongamos ahora que (4.14) se cumple para algún $2 \leq n \leq N$. Entonces, en

virtud de (4.4),

$$V_{n-1}^{m(N)} = \text{mín}\left\{\lambda_0 f_{\theta_0}^{m(n-1)}, \lambda_1 f_{\theta_1}^{m(n-1)}, c_n(m_n)\sum_{i=1}^{K}\gamma_i f_{\vartheta_i}^{m(n-1)} + \mathcal{I}_{m_n} V_n^{m(N)}\right\} \quad (4.15)$$

Por la hipótesis de inducción,

$$\begin{aligned}
\mathcal{I}_{m_n} V_n^{m(N)} &= \int\cdots\int U_n^{m(N)}\left(\sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij}\right)\prod_{i=1}^{n}\prod_{j=1}^{m_i} h(y_{ij})\, dy_{n1}\ldots dy_{nm_n} \\
&= \prod_{i=1}^{n-1}\prod_{j=1}^{m_i} h(y_{ij}) \int\cdots\int U_n^{m(N)}\left(\sum_{i=1}^{n-1}\sum_{j=1}^{m_i} y_{ij} + \sum_{j=1}^{m_n} y_{nj}\right)\prod_{j=1}^{m_n} h(y_{nj})\, dy_{n1}\ldots dy_{nm_n} \\
&= \mathcal{J}_{m_n} U_n^{m(N)}\left(\sum_{i=1}^{n-1}\sum_{j=1}^{m_i} y_{ij}\right)\prod_{i=1}^{n-1}\prod_{j=1}^{m_i} h(y_{ij}). \qquad (4.16)
\end{aligned}$$

Teniendo en cuenta (4.10), vemos que $\prod_{i=1}^{n-1}\prod_{j=1}^{m_i} h(y_{ij})$ es un factor común en (4.15), por lo tanto puede factorizarse. Denotando $S_{n-1} = \sum_{i=1}^{n-1}\sum_{j=1}^{m_i} y_{ij}$, obtenemos:

$$\begin{aligned}
V_{n-1}^{m(N)} &= \text{mín}\Big\{\lambda_0 g_{\theta_0}^{m(n-1)}(S_{n-1}), \lambda_1 g_{\theta_1}^{m(n-1)}(S_{n-1}), \\
&\qquad c_n(m_n)\sum_{i=1}^{K}\gamma_i g_{\vartheta_i}^{m(n-1)}(S_{n-1}) + \mathcal{J}_{m_n} U_n^{m(N)}(S_{n-1})\Big\}\prod_{i=1}^{n-1}\prod_{j=1}^{m_i} h(y_{ij}) \\
&= U_{n-1}^{m(N)}(S_{n-1})\prod_{i=1}^{n-1}\prod_{j=1}^{m_i} h(y_{ij}),
\end{aligned}$$

lo que completa la inducción. □

Por la definición (4.10), el término $\prod_{i=1}^{n}\prod_{j=1}^{m_i} h(y_{ij})$ es un factor común en cada sumando de la desigualdad en (4.7), incluyendo en el operador del extremo derecho, donde, en virtud de la Proposición 4.1.1, se tiene:

$$\mathcal{I}_{m_{n+1}}(V_{n+1}^{m(N)}(y_1^{(m_1)},\ldots,y_{n+1}^{(m_{n+1})})) = \mathcal{I}_{m_{n+1}}\left(U_{n+1}^{m(N)}\left(\sum_{i=1}^{n+1}\sum_{j=1}^{m_i} y_{ij}\right)\prod_{i=1}^{n+1}\prod_{j=1}^{m_i} h(y_{ij})\right),$$

$$= \mathcal{J}_{m_{n+1}} U_{n+1}^{m(N)}\left(\sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij}\right)\prod_{i=1}^{n}\prod_{j=1}^{m_i} h(y_{ij}).$$

Por lo tanto, el término $\prod_{i=1}^{n}\prod_{j=1}^{m_i} h(y_{ij})$ puede cancelarse en todos los sumandos de (4.7).

De este modo, la expresión para la regla óptima de paro (4.7) resulta ser equivalente a

$$\psi_n(s) = I\left\{\text{mín}\left\{\lambda_0 g_{\theta_0}^{m(n)}(s), \lambda_1 g_{\theta_1}^{m(n)}(s)\right\} \leq c_{n+1}(m_{n+1}) \sum_{i=1}^{K} \gamma_i g_{\vartheta_i}^{m(n)}(s) + \mathcal{J}_{m_{n+1}} U_{n+1}^{m(N)}(s)\right\} \tag{4.17}$$

donde $s = S_n = \sum_{i=1}^{n} \sum_{j=1}^{m_i} y_{ij}$.

De manera análoga, la regla de decisión en (4.8) es equivalente a

$$\phi_n = \phi_n(S_n) = I\left\{\lambda_0 g_{\theta_0}^{m(n)}(S_n) \leq \lambda_1 g_{\theta_1}^{m(n)}(S_n)\right\}. \tag{4.18}$$

Así, en (4.17) y (4.18) tenemos las reglas óptimas para una familia exponencial continua de distribuciones.

De manera muy similar, se obtienen las pruebas secuenciales óptimas por grupos para familias exponenciales discretas, suponiendo que la función de masa de probabilidad tiene la siguiente forma:

$$f_\theta(x) = \exp(\theta x - b(\theta))h(x), \quad x = 0, 1, 2, \ldots \tag{4.19}$$

donde $h(x)$ es una función no negativa y $\sum_{x=0}^{\infty} h(x) = 1$.

En este caso, nuevamente, las reglas óptimas tienen la forma de (4.17)–(4.18), con las debidas correcciones de los elementos involucrados: aplicando (4.12) y (4.13) únicamente a argumentos enteros no negativos, y usando

$$\mathcal{J}_m U(s) = \sum_{k=0}^{\infty} U(s+k)\, h^{*m}(k), \quad s = 0, 1, 2, \ldots$$

en lugar de (4.11).

Concluyamos esta sección formulando los siguientes resultados simples pero importantes, que relacionan los problemas de minimización con y sin restricciones, y que serán utilizados en las aplicaciones de pruebas secuenciales óptimas por grupos que presentaremos a continuación.

Proposición 4.1.2 *Sea (ψ, ϕ) la prueba definida por (4.17) y (4.18) para ciertos λ_0, λ_1, γ, ϑ y tamaños de grupo $m_1, m_2, \ldots, m_N$.*

Entonces, para cualquier otra prueba (ψ', ϕ') tal que

$$\alpha(\psi', \phi') \leq \alpha(\psi, \phi), \quad \beta(\psi', \phi') \leq \beta(\psi, \phi),$$

se cumple que

$$K(\vartheta, \gamma; \psi') \geq K(\vartheta, \gamma; \psi).$$

Esto es una consecuencia directa del método de Lagrange, ya que la prueba definida por (4.17) y (4.18) está construida para minimizar la función de Lagrange (4.2).

Si se desea optimizar aún más el procedimiento de prueba, la siguiente generalización simple puede ser útil. Sea $\mathcal{T}$ una cierta clase de pruebas.

Proposición 4.1.3 *Supongamos que existe una prueba* $(\psi, \phi) \in \mathcal{T}$ *tal que, para toda* $(\psi', \phi') \in \mathcal{T}$*, se cumple*

$$K(\vartheta, \gamma; \psi) + \lambda_0 \alpha(\psi, \phi) + \lambda_1 \beta(\psi, \phi) \leq K(\vartheta, \gamma; \psi') + \lambda_0 \alpha(\psi', \phi') + \lambda_1 \beta(\psi', \phi'), \quad (4.20)$$

para algunas constantes λ_0 *y* λ_1*.*

Entonces, para cualquier prueba $(\psi', \phi') \in \mathcal{T}$ *tal que*

$$\alpha(\psi', \phi') \leq \alpha(\psi, \phi), \quad \beta(\psi', \phi') \leq \beta(\psi, \phi),$$

se cumple que

$$K(\vartheta, \gamma; \psi') \geq K(\vartheta, \gamma; \psi).$$

Ambas proposiciones son casi evidentes.

Si $\mathcal{T}$ consiste en una sola prueba secuencial por grupos, la Proposición 4.1.3 es equivalente a la Proposición 4.1.2. Por ejemplo, esto puede aplicarse a pruebas completamente secuenciales (aquellas con $m_1 = m_2 = \cdots = 1$) truncadas en el paso N, interesantes como aproximaciones (cuando $N \to \infty$) a las pruebas con horizonte infinito, las cuales corresponden a las pruebas secuenciales clásicas de razón de probabilidades (SPRTs, Wald and Wolfowitz, 1948).

No obstante, en lo que sigue, queremos utilizar la Proposición 4.1.3 para clases más generales de pruebas secuenciales por grupos.

De la Proposición 4.1.3 se deduce fácilmente que, para encontrar pruebas óptimas dentro de una clase $\mathcal{T}$, es suficiente encontrar una prueba $(\psi, \phi) \in \mathcal{T}$ que minimice la función de Lagrange dentro de dicha clase. Entonces, su costo esperado ponderado $K(\vartheta, \gamma; \psi)$ será mínimo entre todas las pruebas de $\mathcal{T}$ cuyas probabilidades de error de tipo I y tipo II no excedan, respectivamente, $\alpha(\psi, \phi)$ y $\beta(\psi, \phi)$. Como es habitual, los multiplicadores de Lagrange (λ_0 y λ_1) pueden utilizarse para que dichas probabilidades coincidan con algunos valores específicos α y β, si fuera necesario.

Dado que acompañamos nuestro método con un conjunto de programas computacionales que implementan todos los cálculos necesarios, la elección de los multiplicadores de Lagrange suele ser una tarea sencilla. Esto se debe a que se puede realizar una selección adecuada de λ_0 y λ_1 mediante algunas rondas de prueba y error utilizando el software, y el proceso se ve facilitado por el hecho de que, típicamente, valores mayores de λ_0 reducen $\alpha(\psi, \phi)$ sin afectar mucho a $\beta(\psi, \phi)$, y, de

forma análoga, valores mayores de λ_1 disminuyen $\beta(\psi, \phi)$ sin afectar de manera significativa a $\alpha(\psi, \phi)$.

Dado que el software está diseñado tanto para calcular la prueba óptima como para evaluar sus características numéricas, como las probabilidades de error y el costo muestral esperado, podemos determinar fácilmente los valores apropiados de los multiplicadores analizando cómo responden dichas características al variar los parámetros. Asimismo, existe una forma de analizar el efecto de otros parámetros de entrada, como los que definen el cálculo y/o ponderación de los costos de muestreo, en el momento de planificar el procedimiento de prueba.

4.2. Algoritmos para el cálculo de planes de muestreo óptimos

En esta sección describimos esquemas algorítmicos para calcular pruebas secuenciales óptimas por grupos, así como sus características numéricas, en el caso en que las observaciones provienen de distribuciones pertenecientes a familias exponenciales.

4.2.1. Cálculo de pruebas secuenciales óptimas por grupos

En esta sección se propone un esquema algorítmico para calcular pruebas secuenciales óptimas por grupos, en el caso de familias exponenciales. Las fórmulas exactas se encuentran en la sección 4.1.2; aquí únicamente proponemos un esquema numérico para su cálculo.

La parte más laboriosa es el cálculo de la regla óptima de paro, que se obtiene mediante inducción hacia atrás. En el paso n, la regla óptima de paro (4.17) prescribe detenerse cuando

$$\text{mín}\left\{\lambda_0 g_{\theta_0}^{m(n)}(s), \lambda_1 g_{\theta_1}^{m(n)}(s)\right\} \leq c_{n+1}(m_{n+1}) \sum_{i=1}^{K} \gamma_i g_{\vartheta_i}^{m(n)}(s) + \mathcal{J}_{m_{n+1}} U_n^{m(N)}(s). \quad (4.21)$$

Para resolver numéricamente (4.21), es necesario contar con una rutina computacional para el cálculo de $U_n^{m(N)}(s)$. Un análisis simple de la recurrencia en (4.13) sugiere que el principal desafío consiste en evaluar el operador $\mathcal{J}_m$. Siguiendo a Novikov et al. (2023), proponemos una evaluación aproximada de esta integral usando cuadraturas basadas en los valores en malla (grid) de $U_n^{m(N)}$, almacenados en cada paso sucesivo. Esto solo aplica al caso de familias continuas, ya que en el caso discreto el operador se calcula mediante sumas, y no integrales.

En el caso continuo, existe además una complicación adicional relacionada con la forma desconocida de la región de continuación. Existen ciertos casos en los que se puede demostrar teóricamente que dicha región tiene la forma de un intervalo en cada paso. Este es el caso, por ejemplo, cuando $K = 2$ y $\vartheta_1 = \theta_0$, $\vartheta_2 = \theta_1$ (formulación bayesiana), o cuando $\theta_0 \leq \vartheta_1 \leq \theta_1$ y $K = 1$ (problema modificado de Kiefer–Weiss). Incluso si $K = 1$ pero $\vartheta_1 < \theta_0$, por ejemplo, no existe un resultado general que garantice que la región de continuación tenga forma de intervalo; de hecho, es fácil encontrar ejemplos donde esto no ocurre, incluso en el caso más simple del modelo de Bernoulli. Como en este trabajo consideramos un caso mucho más general, con ϑ_i, γ_i, $i = 1, \ldots, K$ arbitrarios, y también K arbitrario, no existe ninguna garantía de que la región de continuación de la prueba óptima tenga siempre la forma de un intervalo.

No obstante, proponemos utilizar la forma de intervalos para las regiones de continuación, pensando que en muchos casos esta forma puede ofrecer buenos resultados, incluso si no son exactamente óptimos. En principio, siempre existe la posibilidad de intentar mejorar el resultado numérico obtenido con una única región de continuación tipo intervalo, permitiendo dos o más intervalos dentro de dicha región. Esto requiere algo de programación adicional, pero puede justificarse si el problema aplicado concreto lo amerita. Los problemas de diseño óptimo con familias exponenciales discretas no presentan esta complicación, ya que las regiones de continuación pueden tratarse punto por punto, en lugar de por intervalos.

Pasemos ahora al cálculo numérico de las reglas óptimas (4.17) y (4.18) para familias exponenciales de distribuciones. En otras palabras, debemos resolver numéricamente la desigualdad (4.21) para $n = N - 1, \ldots, 1$, y

$$\lambda_0 g_{\theta_0}^{m(n)}(s) \leq \lambda_1 g_{\theta_1}^{m(n)}(s) \tag{4.22}$$

para $n = N, \ldots, 1$.

Resolviendo (4.22) con respecto a s, se obtiene fácilmente que la regla de decisión es

$$\phi_n = I_{\{S_n \geq l_n\}}, \quad \text{donde } l_n = \frac{\log(\lambda_0/\lambda_1) + (b(\theta_1) - b(\theta_0)) \sum_{k=1}^{n} m_k}{\theta_1 - \theta_0} \tag{4.23}$$

con $S_n = \sum_{i=1}^{n} \sum_{j=1}^{m_i} y_{ij}$ para cada etapa $n = 1, 2, \ldots, N$ (suponiendo que $\theta_0 < \theta_1$).

Describamos, paso a paso, el algoritmo para la resolución numérica de (4.21), utilizando inducción hacia atrás desde $n = N$ hasta $n = 1$.

Por definición, $\psi_N \equiv 1$. Descendiendo al paso $n = N - 1$, obtenemos los valores

a_{N-1} y b_{N-1} resolviendo (con respecto a s) la ecuación

$$\text{mín}\left\{\lambda_0 g_{\theta_0}^{m(N-1)}(s), \lambda_1 g_{\theta_1}^{m(N-1)}(s)\right\} = c_N(m_N)\sum_{i=1}^{K}\gamma_i g_{\vartheta_i}^{m(N-1)}(s) + \mathcal{J}_{m_N}U_N^{m(N)}(s), \tag{4.24}$$

donde, según la definición (4.12) y la fórmula (4.23), se tiene

$$U_N^{m(N)}(s) = \begin{cases}\lambda_0 g_{\theta_0}^{m(N)}, & \text{si } s > l_N\\ \lambda_1 g_{\theta_1}^{m(N)}, & \text{en caso contrario.}\end{cases}$$

Así, podemos dividir la integral $\mathcal{J}_{m_N}U_N^{m(N)}(s) = \int_{\mathbb{R}} U_N^{m(N)}(s+x)\,h^{*m_N}(x)\,dx$ en dos partes:

$$\mathcal{J}_{m_N}U_N^{m(N)}(s) = \int_{-\infty}^{l_N}\lambda_1 g_{\theta_1}^{m(N)}(s+x)\,h^{*m_N}(x)\,dx + \int_{l_N}^{+\infty}\lambda_0 g_{\theta_0}^{m(N)}(s+x)\,h^{*m_N}(x)\,dx. \tag{4.25}$$

Es fácil ver que

$$\int_{-\infty}^{l_N} g_{\theta_1}^{m(N)}(s+x)h^{*m_N}(x)dx = g_{\theta_1}^{m(N-1)}(s)P_{\theta_1}(\sum_{j=1}^{m_N}Y_{Nj} \le l_N - s) \tag{4.26}$$

$$= g_{\theta_1}^{m(N-1)}(s)P_{\theta_1}^{(m_N)}(l_N - s) \tag{4.27}$$

donde

$$P_\theta^{(m)}(y) = P_\theta\left(\sum_{j=1}^{m}Y_{1j} \le y\right). \tag{4.28}$$

De manera análoga, la segunda integral en (4.25) es

$$\int_{l_N}^{+\infty} g_{\theta_0}^{m(N)}(s+x)\,h^{*m_N}(x)\,dx = g_{\theta_0}^{m(N-1)}(s)\left(1 - P_{\theta_0}^{(m_N)}(l_N - s)\right).$$

Normalmente, la evaluación numérica de la función de distribución en (4.28) en R no representa ninguna dificultad (en muchos casos puede incluso calcularse analíticamente).

Ahora que disponemos de un método para calcular el operador $\mathcal{J}_{m_N}U_N^{m(N)}$ en (4.13), determinamos numéricamente los valores a_{N-1} y b_{N-1} como las raíces de la ecuación (4.24) para $n = N$, y almacenamos los valores de $U_{N-1}^{m(N)}$ evaluados en los puntos de una malla sobre el intervalo (a_{N-1}, b_{N-1}).

Como se explicó al inicio de esta sección, buscamos que la región de continuación tenga la forma de un intervalo (a_n, b_n) para $n = 1, 2, \ldots, N-1$. A pesar de que la región de continuación óptima podría tener una forma más compleja, segui-

mos actuando como si existieran únicamente dos raíces de la ecuación (4.24) (y de las correspondientes en los pasos siguientes hacia abajo): tomamos como región de continuación el intervalo cuyos extremos son esas dos raíces.

Supongamos ahora que ya disponemos del intervalo (a_n, b_n) y de los valores (aproximados) $\tilde{U}_n^{m(N)}(s)$ de $U_n^{m(N)}(s)$ en los puntos $s \in (a_n, b_n)$ de una malla, y que queremos obtener (a_{n-1}, b_{n-1}) y los valores $\tilde{U}_{n-1}^{m(N)}(s)$ para $s \in (a_{n-1}, b_{n-1})$.

Resolviendo con respecto a s la ecuación

$$\text{mín}\left\{\lambda_0 g_{\theta_0}^{m(n-1)}(s), \lambda_1 g_{\theta_1}^{m(n-1)}(s)\right\} = c_n(m_n)\sum_{i=1}^{K}\gamma_i g_{\vartheta_i}^{m(n-1)}(s) + \mathcal{J}_{m_n}\tilde{U}_n^{m(N)}(s), \quad (4.29)$$

obtendremos el intervalo (a_{n-1}, b_{n-1}) y los valores $\tilde{U}_{n-1}^{m(N)}(s)$ para $s \in (a_{n-1}, b_{n-1})$.

Por lo tanto, necesitamos una forma de evaluar (aproximadamente) el operador

$$\mathcal{J}_{m_n}\tilde{U}_n^{m(N)}(s) = \int_{\mathbb{R}} \tilde{U}_n^{m(N)}(s+x)\, h^{*m_n}(x)\, dx. \quad (4.30)$$

Dado que

$$\mathcal{J}_{m_n}\tilde{U}_n^{m(N)}(s) = \int_{-\infty}^{a_n - s} \lambda_1 g_{\theta_1}^{m(n)}(s+x)\, h^{*m_n}(x)\, dx + \int_{a_n - s}^{b_n - s} \tilde{U}_n^{m(N)}(s+x)\, h^{*m_n}(x)\, dx + \int_{b_n - s}^{+\infty} \lambda_0 g_{\theta_0}^{m(n)}(s+x)\, h^{*m_n}(x)\, dx, \quad (4.31)$$

tenemos, para la primera y la tercera integral (como en (4.25))

$$\int_{-\infty}^{a_n - s} g_{\theta_1}^{m(n)}(s+x)\, h^{*m_n}(x)\, dx = g_{\theta_1}^{m(n-1)}(s)\, P_{\theta_1}^{(m_n)}(a_n - s), \quad (4.32)$$

$$\int_{b_n - s}^{+\infty} g_{\theta_0}^{m(n)}(s+x)\, h^{*m_n}(x)\, dx = g_{\theta_0}^{m(n-1)}(s)\, \left(1 - P_{\theta_0}^{(m_n)}(b_n - s)\right). \quad (4.33)$$

Y la integral intermedia en (4.31) puede calcularse utilizando cualquier cuadratura conveniente, basada en los valores previamente almacenados de $\tilde{U}_n^{m(N)}(s)$ en los puntos de la malla dentro del intervalo (a_n, b_n).

Aplicando ahora los mismos pasos a la ecuación (4.29), con $n-2$ en lugar de $n-1$, etc., obtenemos como resultado los valores aproximados de $(a_1, b_1), \ldots, (a_{N-1}, b_{N-1})$ que definen la regla óptima de detención. La regla óptima de decisión ϕ_n siempre viene dada por (4.23).

Cabe señalar que este algoritmo falla en el caso de que la ecuación (4.29) tenga una sola raíz o ninguna. Esto puede ocurrir, por ejemplo, en el contexto de Kiefer-Weiss cuando $\theta_0 < \vartheta_1 < \theta_1$ y todos los grupos son de tamaño 1, y n es demasiado

grande, simplemente porque la prueba secuencial completamente óptima tiene un tiempo de detención acotado (ver, por ejemplo, Lorden, 1980); en consecuencia, no se puede encontrar una región de continuación para valores tan grandes de n. En nuestra implementación del algoritmo en R (véase la sección 4 más adelante), no se contempla este tipo de error, por lo que se recomienda usar grupos de menor tamaño y/o un menor número de grupos si se sospecha que "no existe una región de continuación".

4.2.2. Característica de operación, costo muestral esperado y fórmulas relacionadas

En esta subsección se obtienen fórmulas para calcular las probabilidades de error, el costo muestral esperado, y algunas características relacionadas de las pruebas secuenciales por grupos.

Consideramos pruebas $\langle \psi, \phi \rangle \in \mathscr{S}^N$ tales que $\psi_n = \psi_n(S_n)$ y $\phi_n = \phi_n(S_n)$, para $n = 1, \ldots, N$, donde $S_n = \sum_{i=1}^{n} \sum_{j=1}^{m_i} y_{ij}$.

Denotemos por $D_n^N = D_n^N(\psi, \phi)$ al evento que significa "se acepta H_0 en o después de la etapa n como resultado de aplicar la prueba (ψ, ϕ)". En particular, la característica de operación de una prueba $(\psi, \phi) \in \mathscr{S}^N$ puede calcularse como $OC_\theta(\psi, \phi) = P_\theta(D_1^N(\psi, \phi))$.

Proposición 4.2.1 *Definamos*

$$d_\theta^N(s) = 1 - \phi_N(s)$$

y, recursivamente para $n = N-1, N-2, \ldots, 1$,

$$d_\theta^n(s) = (1 - \phi_n(s))\psi_n(s) + E_\theta d_\theta^{n+1}(s + \sum_{j=1}^{m_n} Y_{nj})(1 - \psi_n(s)) \tag{4.34}$$

Entonces,

$$d_\theta^n(S_n) = P_\theta(D_n^N \mid Y_1^{(m_1)}, \ldots, Y_n^{(m_n)}) \tag{4.35}$$

para $n = 1, 2, \ldots, N$, *y*

$$P_\theta(D_1^N) = OC_\theta(\psi, \phi) = E_\theta d_\theta^1(S_1)$$

Demostración. Por inducción. Para $n = N$:

$$P_\theta(D_N^N \mid Y_1^{(m_1)}, \ldots, Y_N^{(m_N)}) = 1 - \phi_N(S_N) = d_\theta^N(S_N)$$

por la definición de ϕ_N.

Supongamos ahora que (4.35) se cumple para algún $n \leq N$. Entonces:

$$\begin{aligned}
d_\theta^{n-1}(S_{n-1}) &= (1-\phi_{n-1}(S_{n-1}))\psi_{n-1}(S_{n-1}) \\
&\quad + E_\theta\{d_\theta^n(S_n) \mid Y_1^{(m_1)}, \ldots, Y_{n-1}^{(m_{n-1})}\}(1-\psi_{n-1}(S_{n-1})) \\
&= (1-\phi_{n-1}(S_{n-1}))\psi_{n-1}(S_{n-1}) \\
&\quad + E_\theta\{P_\theta(D_n^N \mid Y_1^{(m_1)}, \ldots, Y_n^{(m_n)}) \mid Y_1^{(m_1)}, \ldots, Y_{n-1}^{(m_{n-1})}\}(1-\psi_{n-1}(S_{n-1})) \\
&= (1-\phi_{n-1}(S_{n-1}))\psi_{n-1}(S_{n-1}) \\
&\quad + P_\theta(D_n^N \mid Y_1^{(m_1)}, \ldots, Y_{n-1}^{(m_{n-1})})(1-\psi_{n-1}(S_{n-1})) \\
&= E_\theta\big[(1-\phi_{n-1}(S_{n-1}))\psi_{n-1}(S_{n-1}) \\
&\quad + P_\theta(D_n^N \mid Y_1^{(m_1)}, \ldots, Y_{n-1}^{(m_{n-1})})(1-\psi_{n-1}(S_{n-1})) \mid Y_1^{(m_1)}, \ldots, Y_{n-1}^{(m_{n-1})}\big] \\
&= P_\theta(D_{n-1}^N \mid Y_1^{(m_1)}, \ldots, Y_{n-1}^{(m_{n-1})})
\end{aligned}$$

Por lo tanto, (4.35) también se cumple para $n-1$. □

Sea τ_ψ el tiempo de paro generado por la regla de paro ψ.

Proposición 4.2.2 *Sea k un entero tal que $k \leq N-1$. Definamos*

$$t_{\theta,k}^k(s) = 1 - \psi_k(s)$$

y, recursivamente para $n = k, k-1, \ldots, 2$,

$$t_{\theta,n-1}^k(s) = E_\theta t_{\theta,n}^k\left(s + \sum_{j=1}^{m_n} Y_{nj}\right)(1-\psi_{n-1}(s)). \tag{4.36}$$

Entonces, para todo n entre 1 y k se cumple

$$t_{\theta,n}^k(S_n) = E_\theta\{(1-\psi_n)\ldots(1-\psi_k) \mid Y_1^{(m_1)}, \ldots, Y_n^{(m_n)}\}. \tag{4.37}$$

En particular,

$$P_\theta(\tau_\psi > k) = E_\theta t_{\theta,1}^k\left(\sum_{j=1}^{m_1} Y_{1j}\right).$$

Demostración. Por inducción sobre $n = k, k-1, \ldots, 2$.

Para $n = k$:

$$t_{\theta,k}^k(S_k) = 1 - \psi_k(S_k) = E_\theta\{1-\psi_k(S_k) \mid Y_1^{(m_1)}, \ldots, Y_k^{(m_k)}\}$$

por la definición de ψ.

Supongamos que (4.37) se cumple para algún $2 \leq n \leq k$, es decir,

$$t_{\theta,n}^k(S_n) = E_\theta\{(1-\psi_n)\ldots(1-\psi_k) \mid Y_1^{(m_1)}, \ldots, Y_n^{(m_n)}\}.$$

Entonces:

$$t^k_{\theta,n-1}(S_{n-1}) = E_\theta\{t^k_{\theta,n}(S_{n-1} + \sum_{j=1}^{m_n} Y_{nj}) \mid Y_1^{(m_1)}, \ldots, Y_{n-1}^{(m_{n-1})}\}(1 - \psi_{n-1}(S_{n-1})) =$$

$$= E_\theta\{t^k_{\theta,n}(\sum_{i=1}^{n}\sum_{j=1}^{m_i} Y_{ij}) \mid Y_1^{(m_1)}, \ldots, Y_{n-1}^{(m_{n-1})}\}(1 - \psi_{n-1}(S_{n-1})) =$$

$$= E_\theta\{E_\theta\{(1-\psi_n)\ldots(1-\psi_k) \mid Y_1^{(m_1)}, \ldots, Y_n^{(m_n)}\} \mid Y_1^{(m_1)}, \ldots, Y_{n-1}^{(m_{n-1})}\}(1-\psi_{n-1}(S_{n-1})) =$$

$$= E_\theta\{(1 - \psi_n)\ldots(1 - \psi_k) \mid Y_1^{(m_1)}, \ldots, Y_{n-1}^{(m_{n-1})}\}(1 - \psi_{n-1}(S_{n-1})) =$$

$$= E_\theta\{(1 - \psi_{n-1})\ldots(1 - \psi_k) \mid Y_1^{(m_1)}, \ldots, Y_{n-1}^{(m_{n-1})}\},$$

lo cual demuestra que (4.37) también se cumple para $n-1$. □

Usando la Proposición 4.2.2, podemos calcular el número esperado de grupos como

$$1 + \sum_{n=1}^{N-1} P_\theta(\tau_\psi > n),$$

el número esperado de observaciones como

$$m_1 + \sum_{n=1}^{N-1} m_{n+1} P_\theta(\tau_\psi > n),$$

o bien, el costo muestral esperado $C(\theta;\psi)$ como

$$c_1(m_1) + \sum_{n=1}^{N-1} c_{n+1}(m_{n+1}) P_\theta(\tau_\psi > n). \tag{4.38}$$

Una forma más directa de calcular estas características es incorporar la suma de (4.38) dentro de la recursión dada en (4.36):

Proposición 4.2.3 *Sea*

$$q^N_{\theta,N}(s) = 0$$

y, recursivamente para $n = N, N-1, \ldots, 2$,

$$q^N_{\theta,n-1}(s) = \left(c_n(m_n) + E_\theta q^N_{\theta,n}\left(s + \sum_{j=1}^{m_n} Y_{nj}\right)\right)(1 - \psi_{n-1}(s)) \tag{4.39}$$

Entonces,

$$C(\theta;\psi) = c_1(m_1) + E_\theta q^N_{\theta,1}\left(\sum_{j=1}^{m_1} Y_{1j}\right).$$

Demostración. Demostraremos que, para todo $n < N$,

$$q^N_{\theta,n}(s) = \sum_{k=n}^{N-1} c_k(m_{k+1}) t^k_{\theta,n}(s). \tag{4.40}$$

Por inducción. Para $n = N-1$:

$$q^N_{\theta,N-1}(s) = c_N(m_N)(1-\psi_{N-1}(s)) = c_N(m_N) t^{N-1}_{\theta,N-1}(s).$$

Supongamos ahora que (4.40) se cumple para algún $n < N$. Entonces:

$$\begin{aligned}
q^N_{\theta,n-1}(s) &= (c_n(m_n) + E_\theta q^N_{\theta,n}(s+X_n))(1-\psi_{n-1}(s)) = \\
&= \left(c_n(m_n) + E_\theta \left\{ \sum_{k=n}^{N-1} c_{k+1}(m_{k+1}) t^k_{\theta,n}(s+X_n) \right\} \right)(1-\psi_{n-1}(s)) = \\
&= c_n(m_n)(1-\psi_{n-1}(s)) + \sum_{k=n}^{N-1} c_{k+1}(m_{k+1}) E_\theta t^k_{\theta,n}(s+X_n)(1-\psi_{n-1}(s)) = \\
&= c_n(m_n) t^{n-1}_{\theta,n-1}(s) + \sum_{k=n}^{N-1} c_{k+1}(m_{k+1}) t^k_{\theta,n-1}(s) = \sum_{k=n-1}^{N-1} c_{k+1}(m_{k+1}) t^k_{\theta,n-1}(s).
\end{aligned}$$

Y entonces, usando la Proposición 4.2.2:

$$\begin{aligned}
c_1(m_1) + E_\theta q^N_{\theta,1}(X_1) &= c_1(m_1) + E_\theta \sum_{k=1}^{N-1} c_{k+1}(m_{k+1}) t^k_{\theta,1}(X_1) = \\
&= c_1(m_1) + \sum_{k=1}^{N-1} c_{k+1}(m_{k+1}) P_\theta(\tau_\psi > k) = \sum_{k=1}^{N} c_k(m_k) P_\theta(\tau_\psi \geq k)
\end{aligned}$$

Recordando que $P_\theta(\tau_\psi \geq k) = \sum_{n=k}^N P_\theta(\tau_\psi = n)$, reescribimos la expresión anterior como suma doble:

$$c_1(m_1) + E_\theta q^N_{\theta,1}(X_1) = \sum_{k=1}^{N} \sum_{n=k}^{N} c_k(m_k) P_\theta(\tau_\psi = n)$$

Agrupando los términos de la forma $P_\theta(\tau_\psi = n)$, obtenemos:

$$c_1(m_1) + E_\theta q^N_{\theta,1}(X_1) = \sum_{n=1}^{N} \sum_{k=1}^{n} c_k(m_k) P_\theta(\tau_\psi = n) = C(\theta;\psi).$$

□

Si la región de continuación de la regla de paro ψ en las Proposiciones 4.2.1, 4.2.2,

4.2.3 tiene la forma de un intervalo, entonces, en el caso de familias exponenciales continuas, podemos utilizar, para las fórmulas recurrentes (4.34), (4.36) y (4.39), una modificación obvia del algoritmo propuesto en la Sección 4.2.1 para el cálculo de las pruebas óptimas.

En el caso discreto, nuevamente, las fórmulas recurrentes correspondientes se basan en sumas en lugar de integrales, por lo tanto, los puntos de la región de continuación pueden almacenarse, en los pasos sucesivos de la recurrencia, de forma individual en lugar de por intervalos, y no se requiere ninguna suposición previa sobre la forma de dicha región.

4.3. Software complementario

Esta sección describe el código en `R` que desarrollamos como implementación de los algoritmos de la Sección 4.2, para los casos particulares de las familias de distribuciones binomial y normal. El código está disponible en el repositorio público de GitHub: `https://github.com/tosinabase/groupsequential`. Todos los resultados numéricos presentados en este artículo se obtuvieron utilizando este código, y pueden reproducirse mediante llamadas simples a las funciones que ya están disponibles en el repositorio. Además, se proporciona un ejemplo de uso para el cálculo de una prueba bayesiana en dos etapas para la media normal y sus características de desempeño.

El repositorio está estructurado en dos carpetas principales, correspondientes a los modelos binomial y normal. Cada carpeta contiene scripts en `R` para el cálculo de planes óptimos y sus características de desempeño, junto con ejemplos de uso respectivos. Información detallada sobre la estructura de carpetas, el contenido de los archivos y las instrucciones de uso se encuentra en el archivo `README.md` incluido en el repositorio.

Para ilustrar el uso de nuestro código, consideremos un ejemplo sencillo del cálculo de un plan de prueba en dos etapas y de sus características de desempeño, en el contexto de las pruebas de hipótesis sobre la media normal presentado en Tarima and Flournoy (2022) (Ejemplo 2). Se supone que en la primera etapa se pueden tomar $n_1 = 100$ observaciones, y en la segunda, $n_2 = 100$. Todas las observaciones se suponen independientes dentro de cada grupo y entre los grupos, y que siguen una distribución normal $N(\theta, 1)$. El valor medio es $\theta = \theta_0 = 0$ bajo la hipótesis H_0, y $\theta = \theta_1 = 0.21$ bajo H_1.

En el Listado 4.1 se muestran los parámetros de entrada para este ejemplo, bajo un enfoque bayesiano con $\gamma_1 = \gamma_2 = 0.5$, $\vartheta_1 = \theta_0$, $\vartheta_2 = \theta_1$ y $c_n(m) = m$. Los valores de los multiplicadores de Lagrange, $\lambda_0 = 1526$ y $\lambda_1 = 326$, fueron seleccionados

y ajustados para alcanzar las probabilidades deseadas de error de tipo I y tipo II: $\alpha = 0.025$ y $\beta = 0.2$ (cf. Tabla 2 en Tarima and Flournoy, 2022).

Listado 4.1: Inicialización y cálculo del plan óptimo

```
# use the repository code
source("./normal/plan.R")
source("./normal/characteristics.R")

# input parameters:
# hypothesized values:
th0 <- 0 # theta0
th1 <- 0.21 # theta1
thgam <- c(th0, th1) # points at which ESC is weighted
gam <- c(0.5, 0.5) # weights for weighting ESC

# Lagrangian multipliers
l0 <- 1526 # lambda0
l1 <- 326 # lambda1

M <- c(100, 100) # group sizes
precision <- 0.05 # grid step
cost_fn <- function (x) x # cost of a group group of x observations

# main function call for optimal plan computation:
test <- calculate_plan(M, l0, l1, th0, th1, gam, thgam,
                       cost_fn=cost_fn,
                       precision=precision,
                       print_output=FALSE)
```

Listado 4.2: Intervalo de continuación para el primer paso

```
print(paste(head(test$data[[1]]$grid, 1), tail(test$data[[1]]$grid, 1)))
```

Result:

```
[1] "8.60693365315347 23.3352164728895"
```

Listado 4.3: Constante crítica para el segundo paso

```
print(test$data[[2]]$const)
```

Result:

```
[1] 28.35004
```

Los resultados muestran que el plan bayesiano óptimo en dos etapas continúa a la segunda etapa si la suma muestral $S_1 \in (8.61, 23.34)$, y en la segunda etapa se rechaza H_0 cuando $S_2 > 28.35$, aceptándose H_0 en caso contrario, donde S_2 es la suma muestral basada en ambas muestras.

El Listado 4.4 muestra el cálculo de las probabilidades de error mediante la llamada a la función de característica de operación.

Listado 4.4: Cálculo de las probabilidades de error de tipo I y tipo II

```
alpha <- 1 - operating_characteristic(test, th0)
beta <- operating_characteristic(test, th1)
print(paste("alpha:", alpha, ", beta:", beta))
```

Resultado:

```
[1] "alpha: 0.0250017679625842 , beta: 0.200056694872385"
```

Se observa que, para la prueba resultante, α es muy cercana a 0.025 y β es muy cercana a 0.2.

Listado 4.5: Cálculo del número promedio de observaciones bajo H_0 y H_1

```
ESC0 <- expected_sampling_cost(test, th0, cost_fn=cost_fn)
ESC1 <- expected_sampling_cost(test, th1, cost_fn=cost_fn)
print(paste("ESC0:", ESC0, ", ESC1:", ESC1))
```

Resultado:

```
[1] "ESC0: 118.489308233436 , ESC1: 148.471027781829"
```

El costo muestral esperado (número esperado de observaciones) de la prueba obtenida es 118.5 bajo H_0, y 148.5 bajo H_1. Como comparación, todos los planes presentados en la Tabla 2 de Tarima and Flournoy (2022) tienen un número esperado de observaciones bajo H_0 de casi 200.

El tamaño de muestra fijo correspondiente a $\alpha = 0.025$, $\beta = 0.2$, $\theta_0 = 0$ y $\theta_1 = 0.21$ es 177.78.

4.4. Aplicaciones y resultados numéricos

En esta sección se presentan los resultados del cálculo numérico de la prueba secuencial por grupos óptima descrita en las Secciones 4.1 y 4.2, así como la evaluación de sus características de desempeño. El contexto de estas aplicaciones se encuentra en la literatura.

4.4.1. Planes de muestreo óptimos para inspección por atributos

En esta sección presentamos una aplicación de nuestro método en el contexto discutido en Hu and Wang (2018b), en relación con los procedimientos estandarizados para planes de muestreo por atributos provistos por ISO2859-1 (1999). La Tabla 2 de Hu and Wang (2018b) reporta el costo muestral esperado que presentan diversos planes de ISO2859-1 (1999) para varios valores de θ_0, θ_1, α, β, y otros parámetros del problema de prueba. En particular, la estructura de costos, en nuestros términos, se asumió como $c_n(m_n) = a + bm_n$ para todas las etapas n. Se consideró un enfoque bayesiano, con $\gamma_1 = \gamma_2 = 0.5$. Se reportaron los resultados para planes con $N = 2$ y $N = 5$ grupos. El número total de planes considerados fue 24.

Para cada uno de estos casos aplicamos nuestro método propuesto, con una optimización adicional sobre los tamaños de grupo (aplicando la Proposición 4.1.3, con $\mathcal{T}$ definida como la clase de todas las pruebas secuenciales por grupos con un número fijo de etapas). La Proposición 4.1.3 muestra que el plan obtenido como resultado de esta optimización tiene un costo total esperado (ponderado) mínimo entre todas las pruebas secuenciales por grupos con las mismas probabilidades de error o menores. Para la optimización numérica, en principio, puede utilizarse cualquier rutina general de minimización. Nos llamó la atención el algoritmo genético implementado por Sekhon and Mebane, W. R. Jr. (2011), el cual funcionó de forma eficiente para este tipo de problemas. Cabe señalar, sin embargo, que el algoritmo genético no garantiza alcanzar siempre el valor mínimo, por lo que las pruebas obtenidas mediante este enfoque no siempre son óptimas en el sentido de la Proposición 4.1.3. No obstante, los resultados obtenidos en los ejemplos numéricos se comparan favorablemente con los de los métodos existentes. Por otro lado, la aplicación de un algoritmo numérico de minimización exacta garantizaría la verdadera optimalidad de los planes obtenidos.

Presentamos nuestros resultados numéricos en la Tabla 4.1. Las columnas ESC_{ISO}, α_{ISO} y β_{ISO} corresponden a los valores originales de la Tabla 2 de Hu and Wang (2018b), y las columnas ESC, α y β a los valores calculados de nuestro plan óptimo. Es fácil ver que los planes que obtuvimos son claramente superiores a los estándares (más precisamente, para la eficiencia definida como $(ESC_{ISO} - ESC)/ESC_{ISO} \times 100$, el primer cuartil es 2.6 %, la mediana es 7.7 %, y el tercer cuartil es 15 %). Este efecto se debe a la optimización de los tamaños de grupo, manteniendo el mismo número de grupos que en el estándar. La mejora total debe atribuirse tanto al carácter óptimo de las pruebas utilizadas como a la optimización de los tamaños de grupo. La Tabla 4.2 contiene los parámetros de las pruebas correspondientes a los resultados reportados en la Tabla 4.1 (λ_0, λ_1 y $m_1, \ldots, m_N$), lo que permite la reproducibilidad

de los resultados.

En Hu and Wang (2018b), se propuso un método numérico para la construcción de pruebas secuenciales por grupos "casi óptimas" para proporciones binomiales, basado en algoritmos de búsqueda directa en lugar de programación dinámica. El algoritmo optimiza internamente entre las fronteras de paro, el número de grupos, los tamaños de grupo, etc. La Tabla 2 de Hu and Wang (2018b) presenta los resultados obtenidos por su algoritmo para cada uno de los ejemplos considerados en el estándar ISO. Los resultados muestran que los planes del estándar ISO tienen un gran potencial de mejora, ya que el mínimo del costo esperado alcanzado por el método de Hu and Wang (2018b) es notablemente inferior al proporcionado por el estándar. Porque que el algoritmo de Hu and Wang (2018b) aún no es óptimo, aplicamos nuestro método (véase la Proposición 4.1.3) a cada uno de los casos considerados en su Tabla 2, utilizando la misma rutina de minimización de Lagrange con optimización del tamaño de grupo de que está escrito anteriormente. Los resultados se presentan en la Tabla 4.3. La columna ESC_{HW} contiene el valor mínimo calculado reportado por Hu and Wang (2018b), para los valores originales de α_{ISO} y β_{ISO}, y las columnas ESC, α y β corresponden a los valores calculados para nuestro plan óptimo. Es fácil ver que los planes obtenidos por nuestro método siguen siendo superiores a los de Hu and Wang (2018b) (más precisamente, para la eficiencia definida como $(ESC_{HW} - ESC)/ESC_{HW} \times 100$, el primer cuartil es 0.26 %, la mediana es 1.02 % y el tercer cuartil es 2.65 %). A pesar de que los resultados de Hu and Wang (2018b) ya están optimizados, nuestro método mejora aún más el costo esperado (ponderado) en el 50 % de los casos por al menos 1.02 %, y en el 25 % de los casos, por al menos 2.65 %, con una mejora máxima superior al 8 %. Este efecto debe atribuirse principalmente al carácter óptimo de las reglas obtenidas mediante la minimización de Lagrange en la que se basan nuestras pruebas (es decir, esencialmente a la inducción hacia atrás), ya que la optimización de los tamaños de grupo ya está incorporada en el algoritmo de Hu and Wang (2018b). Nuevamente, la Tabla 4.4 correspondiente contiene los parámetros de las pruebas óptimas presentadas en la Tabla 4.3.

4.4.2. Planes óptimos en ensayos clínicos de Fase II

Las pruebas secuenciales por grupos se utilizan ampliamente en aplicaciones de ensayos clínicos (véase, por ejemplo, Jennison and Turnbull, 1999). En esta sección aplicamos los algoritmos presentados en la Sección 3 a ensayos clínicos de Fase II.

Nos referimos al contexto de Fleming (1982), donde se propone la construcción de un plan de muestreo secuencial por grupos para ensayos clínicos diseñados para evaluar el efecto terapéutico de tratamientos contra el cáncer, basándose en la fre-

cuencia de "regresiones" tumorales después de aplicar el tratamiento. Se asume que las frecuencias siguen una distribución binomial con parámetro θ, que representa la probabilidad de regresiones. Las hipótesis de interés son $H_0 : \theta \leq \theta_0$ frente a $H_1 : \theta > \theta_1$, donde $\theta_0 < \theta_1$. Como es habitual en las aplicaciones, el intervalo $[\theta_0, \theta_1]$ se considera una zona de indiferencia, y queremos aplicar nuestro diseño óptimo de la Sección 4.1 para probar una hipótesis simple $H_0 : \theta = \theta_0$ frente a $H_1 : \theta = \theta_1$. Para los valores nominales de $\alpha = 0.05$ y $\beta = 0.1$, y para una serie de valores de θ_0 y θ_1, Fleming (1982) proporciona una serie de planes secuenciales por grupos, variando los tamaños m_i de los grupos con el fin de aproximarse a los valores de α y β, y tratando de mantener el número esperado de observaciones ESN_0 y ESN_1 bajo las hipótesis H_0 y H_1 en un nivel razonablemente bajo. Algunas combinaciones seleccionadas de estos parámetros se muestran en la Tabla 12.1 de Jennison and Turnbull (1999). En este ejemplo comparamos nuestros planes con los de dicha tabla. Tomamos $c_i(m) = m$ para todo i y m (de modo que el costo muestral esperado coincide, en este caso, con el número esperado de observaciones, ESN). Además de los casos $N = 2$ y $N = 3$, calculamos nuestros planes para $N = 4$, 5 y 10 grupos. Usamos $\gamma_1 = \gamma_2 = 0.5$ para ponderar el ESN entre $\vartheta_1 = \theta_0$ y $\vartheta_2 = \theta_1$, lo que corresponde al marco bayesiano con probabilidades a priori iguales.

Para obtener mejores aproximaciones a α y β, utilizamos una optimización adicional sobre los tamaños de grupo, exactamente como en el ejemplo de la sección anterior 4.4.1. En virtud de la Proposición 4.1.3, la optimización adicional sobre los tamaños de grupo proporciona valores más bajos de la función de Lagrange, ofreciendo una mejor aproximación a los valores nominales de α y β. Usamos el mismo algoritmo de minimización numérica que en la sección 4.4.1, buscando el mínimo de la función de Lagrange sobre $m_1, m_2, \ldots, m_N$ con un número de grupos dado. Después de esto, se utilizan los multiplicadores λ_0 y λ_1 para cumplir con los requisitos sobre α y β. La naturaleza discreta de las probabilidades binomiales no permite, en términos generales, igualar exactamente las probabilidades de error a α y β, por lo que empleamos aproximaciones sucesivas probando diferentes valores de los multiplicadores para acercarnos a los valores nominales. Este "modo manual" no representa ninguna dificultad, ya que valores más altos de λ_0 hacen que α disminuya, produciendo solo un efecto menor sobre β; el efecto de λ_1 sobre β es bastante simétrico.

Presentamos los resultados numéricos de la comparación con el diseño de Fleming en las Tablas 4.5 a 4.8. Cada tabla corresponde a un par específico de (θ_0, θ_1) y yuxtapone las respectivas características de desempeño para ambos métodos. En este contexto, N denota el número de grupos, y α, β corresponden a las probabilidades de error de tipo I y tipo II, respectivamente. Las tablas reportan el tamaño muestral esperado bajo la hipótesis nula ESS_0, bajo la hipótesis alternativa ESS_1, y el tamaño

muestral máximo MSS $= m_1 + m_2 + \cdots + m_N$, que corresponde al número total de observaciones cuando no se produce paro anticipado. Adicionalmente, se proporciona la razón $t =$ MSS/FSS, donde FSS denota el tamaño muestral fijo correspondiente.

La comparación revela que el diseño propuesto logra un tamaño muestral esperado menor que los planes de Fleming (1982), en diferentes configuraciones. Por ejemplo, en el caso de $\theta_0 = 0.3$, $\theta_1 = 0.5$ y 2 grupos, el tamaño muestral esperado bajo la alternativa se reduce de 40.8 a 38.7. Reducciones similares se obtienen en otros casos. De manera muy natural, aumentar el número de grupos implica una reducción adicional del tamaño muestral esperado.

Es notable que, incluso para pruebas con un número pequeño de grupos, nuestro método proporciona una aproximación muy cercana de las probabilidades de error α y β a sus niveles nominales, con una mejora adicional al aumentar el número de grupos.

Concluimos que nuestro método propuesto puede recomendarse para aplicaciones dentro del modelo de respuesta binaria, al menos cuando el número de grupos no sea muy grande. Para un número muy grande de grupos, se espera que la optimización sobre los tamaños de grupo crezca considerablemente, provocando un aumento en el tiempo de computación. En tal caso, recomendaríamos, como alternativa, considerar planes con grupos de tamaño variable (véase, por ejemplo, Novikov, 2024a), que evitan la minimización en alta dimensión (también conocidos en la literatura como *secuencialmente planificados* (Schmitz, 1993) o *adaptativos* (Dragalin, 2006)).

4.4.3. Planes secuenciales por grupos óptimos para la prueba de medias normales

En esta sección tratamos planes secuenciales por grupos óptimos para la prueba de hipótesis sobre una media normal, con varianza conocida, en el contexto de Eales and Jennison (1992). En este ejemplo consideramos únicamente los entornos bayesiano y de Kiefer–Weiss, por ser los más comunes en las aplicaciones, pero todas las demás funciones objetivo consideradas en Eales and Jennison (1992) pueden tratarse de la misma manera, utilizando el software complementario que desarrollamos para este trabajo (de hecho, también puede realizarse un tratamiento aproximado de F_5 empleando un esquema de cuadratura adecuado para el cálculo de la función objetivo). Asimismo, realizamos una optimización adicional sobre el tamaño de grupo, como en los ejemplos de las secciones 4.4.1 y 4.4.2 anteriores, para mejorar la eficiencia de la prueba. Además, analizamos el efecto del parámetro de costo fijo adicional ("overhead") v cuando el costo se define como $c_n(m) = m + v$ para todo n y todo m.

Consideramos el problema de prueba de hipótesis $H_0 : \theta = -\delta$ frente a $H_1 : \theta = \delta$

sobre el valor medio θ de una distribución normal con varianza conocida (que en este ejemplo se supone igual a 1). Realizamos las evaluaciones para cada combinación del tamaño de grupo ($N = 2, 3, 4, 5, 6, 8, 10$) y del parámetro de costo fijo adicional ($v = 0, 3, 6, 10, 15$), utilizando $\alpha = \beta = 0.05$. Para $\delta = 0.1$ en este ejemplo, el tamaño muestral fijo resulta ser

$$FSS = (z_\alpha + z_\beta)^2/(2\delta)^2 = 270.5543,$$

donde z_α es el cuantil superior $\alpha 100\,\%$ de la distribución normal estándar.

Primero, calculamos las pruebas secuenciales óptimas por grupos aplicando un procedimiento de minimización numérica sobre todos los tamaños de grupo. La práctica común en los ensayos clínicos es utilizar el denominado "nivel de información" en lugar del tamaño de grupo entero. Esto hace que el parámetro "tamaño de grupo" sea continuo, de modo que puede emplearse cualquier algoritmo de optimización general que resulte conveniente para la minimización. Nosotros utilizamos el método libre de gradiente propuesto en Nelder and Mead (1965). Los resultados se presentan en las Tablas 4.9 y 4.10.

La Tabla 4.9 contiene los datos de eficiencia calculados para las pruebas bayesianas ($\vartheta_1 = -\delta$, $\vartheta_2 = \delta$, $\gamma_1 = \gamma_2 = 0.5$). Para cada combinación del número de grupos y del parámetro de costo fijo adicional ("overhead"), la eficiencia se calcula como $ESC/(FSS + \text{overhead})$ expresada en porcentaje (columna "Eficiencia (con optim.)"). La desviación estándar (columna DE) presenta los valores calculados de la desviación estándar del número total de observaciones, dividido por el FSS (expresada en porcentaje). Para analizar el efecto de la optimización del tamaño de grupo, también calculamos la eficiencia de una prueba secuencial por grupos que utiliza el mismo número máximo total de observaciones que la prueba optimizada, y el mismo número de grupos de tamaño igual. Los resultados (en porcentaje respecto al costo de un plan de una sola muestra) se muestran en la columna "Eficiencia (sin optim.)". La última columna ("Efecto de la optimización") contiene el efecto de la optimización del tamaño de grupo calculado como "(la razón entre la eficiencia de la prueba óptima sin optimización del tamaño de grupo y la de la prueba óptima optimizada sobre los tamaños de grupo) menos uno", y expresado en porcentaje.

Del análisis de los datos de la Tabla 4.9 se observa que la eficiencia de las pruebas secuenciales óptimas por grupos optimizadas sobre los tamaños de grupo varía desde 72.1 % para 2 grupos hasta 53.2 % para 10 grupos, cuando no existe costo fijo adicional ($v = 0$). Debe tenerse en cuenta que estos valores representan el porcentaje del costo muestral que requiere la prueba secuencial óptima por grupos, con respecto al costo del plan de una sola etapa correspondiente al FSS de una etapa; por lo tanto, valores altos en la "Eficiencia" corresponden a un mayor costo espera-

do. Para las pruebas sin optimización del tamaño de grupo, los valores respectivos van desde 73 % para 2 grupos hasta 54.8 % para 10 grupos (ligeramente menos eficientes, con una ganancia de optimización de hasta 2.9 % para 10 grupos). El efecto del incremento del costo fijo adicional es claramente visible: un costo fijo adicional elevado hace ligeramente menos atractiva la utilización de planes secuenciales, con o sin optimización del tamaño de grupo (de 53.2 % para costo fijo nulo a 64.3 % para un valor de costo fijo adicional de 15, lo que representa aproximadamente un 5.6 % adicional del FSS total, pero aplicado *en cada grupo*). Para los planes sin optimización del tamaño de grupo, la eficiencia es en gran medida la misma, simplemente con un efecto de optimización del tamaño de grupo menos marcado cuando aumenta el costo fijo adicional. Por último, la desviación estándar del tamaño muestral se mantiene prácticamente igual, tanto para los planes optimizados sobre los tamaños de grupo como para los no optimizados, especialmente para valores altos del costo fijo adicional. En planes sin costo fijo adicional, la desviación estándar tiende a ser ligeramente menor para los planes sin optimización del tamaño de grupo en comparación con aquellos optimizados sobre los tamaños de grupo.

La Tabla 4.10 contiene los datos de eficiencia calculados para el problema de Kiefer–Weiss ($\vartheta_1 = 0$, $\gamma_1 = 1$). Los datos presentados siguen las mismas definiciones y la misma descripción general que en la Tabla 4.9.

Del análisis de los datos de la Tabla 4.10 se observa que la eficiencia de las pruebas secuenciales óptimas por grupos optimizadas sobre los tamaños de grupo varía desde 86.5 % para 2 grupos hasta 74.5 % para 10 grupos, cuando no existe costo fijo adicional ($v = 0$). En la formulación de Kiefer–Weiss, la eficiencia global de los planes secuenciales por grupos es marcadamente inferior a la observada bajo la formulación bayesiana. Para las pruebas sin optimización del tamaño de grupo, los valores respectivos van desde 87 % para 2 grupos hasta 74.9 % para 10 grupos (y se observa un efecto muy pequeño de la optimización del tamaño de grupo).

El efecto del incremento del costo fijo adicional es muy similar al de los planes bayesianos: un costo fijo adicional elevado hace considerablemente menos atractiva la utilización de planes secuenciales, con o sin optimización del tamaño de grupo.

Para los planes sin optimización del tamaño de grupo, la eficiencia es en gran medida la misma, pero el efecto de la optimización del tamaño de grupo es más pronunciado cuando aumenta el costo fijo adicional. No consideramos esto de gran importancia, porque en estos casos el efecto de la planificación secuencial en sí resulta bastante cuestionable: por ejemplo, con 6 grupos y un costo fijo adicional de 15 por grupo, observamos que el plan secuencial óptimo requiere, en promedio, un 85.9 % del tamaño muestral fijo (no secuencial), incluso con la optimización sobre el tamaño de grupo. Dado que aún existe una desviación estándar del tamaño muestral de aproximadamente un 31.6 % del FSS, difícilmente se optaría por una planificación

secuencial en un caso así.

Finalmente, de manera muy similar al caso bayesiano, la desviación estándar del tamaño muestral se mantiene prácticamente igual, con o sin optimización del tamaño de grupo, especialmente cuando el valor del costo fijo adicional es relativamente alto. En planes sin costo fijo adicional, la desviación estándar tiende a ser ligeramente menor para los planes sin optimización del tamaño de grupo en comparación con aquellos optimizados sobre los tamaños de grupo.

En general, las Tablas 4.9 y 4.10 pueden ser útiles al planificar pruebas secuenciales por grupos para la media normal, con o sin costo fijo adicional, tanto en el contexto bayesiano como en el contexto de Kiefer–Weiss, con optimización opcional sobre los tamaños de grupo. Los resultados numéricos son absolutos, en el sentido de que son independientes de la distancia entre las hipótesis y de la varianza de las observaciones (siempre que esta última sea conocida). Otros niveles de probabilidades de error α y β y/o otras estructuras de costos pueden analizarse fácilmente utilizando el software complementario que hemos puesto a disposición pública en Farkhshatov and Novikov (2025) (véase la sección 4.3 para su breve descripción). La misma funcionalidad está disponible en Novikov (2024b) para un contexto de múltiples hipótesis.

No.	θ_0	θ_1	**a/b**	**N**	**ESC**	**ESC**$_{ISO}$	α	α_{ISO}	β	β_{ISO}
1	0.935	0.64	10	2	2.353	2.5019	0.048	0.048	0.115	0.118
2	0.935	0.64	5	2	3.376	3.6139	0.046	0.048	0.118	0.118
3	0.935	0.64	1	2	10.915	12.5097	0.045	0.048	0.116	0.118
4	0.935	0.64	0.5	2	23.433	23.6294	0.048	0.048	0.1	0.118
5	0.935	0.64	10	5	2.306	3.3909	0.032	0.048	0.086	0.1353
6	0.935	0.64	5	5	3.408	4.1734	0.031	0.048	0.091	0.1353
7	0.935	0.64	1	5	10.164	10.4335	0.048	0.048	0.134	0.1353
8	0.935	0.64	0.5	5	16.691	18.2586	0.048	0.048	0.134	0.1353
9	0.935	0.696	10	2	2.597	2.6348	0.0504	0.0557	0.096	0.1
10	0.935	0.696	5	2	4.010	4.1241	0.0487	0.0557	0.100	0.1
11	0.935	0.696	1	2	15.529	16.0382	0.0476	0.0557	0.096	0.1
12	0.935	0.696	0.5	2	29.632	30.9308	0.0476	0.0557	0.096	0.1
13	0.935	0.696	10	5	3.185	3.9336	0.0159	0.0374	0.057	0.1
14	0.935	0.696	5	5	4.308	5.2449	0.0367	0.0374	0.082	0.1
15	0.935	0.696	1	5	14.480	15.7346	0.0310	0.0374	0.100	0.1
16	0.935	0.696	0.5	5	28.628	28.8467	0.0279	0.0374	0.088	0.1
17	0.975	0.871	10	2	4.766	4.7797	0.0513	0.0519	0.098	0.1
18	0.975	0.871	5	2	8.386	8.4213	0.0477	0.0519	0.096	0.1
19	0.975	0.871	1	2	36.857	37.5544	0.0507	0.0519	0.097	0.1
20	0.975	0.871	0.5	2	72.193	73.9708	0.0490	0.0519	0.092	0.1
21	0.975	0.871	10	5	5.365	5.812	0.0295	0.0416	0.064	0.1
22	0.975	0.871	5	5	8.421	9.0971	0.0415	0.0416	0.100	0.1
23	0.975	0.871	1	5	32.649	35.3776	0.0415	0.0416	0.100	0.1
24	0.975	0.871	0.5	5	66.516	68.2282	0.0415	0.0416	0.100	0.1

Tabla 4.1: Comparación con la norma ISO 2859-1

No.	λ_0	λ_1	**Tamaños de grupos**				
1	6.0	5.0	5	21			
2	6.0	5.0	5	20			
3	62.0	40.0	7	6			
4	297.0	599.0	10	3			
5	5.0	4.0	10	7	10	11	9
6	12.0	8.0	10	6	6	10	9
7	100.0	84.0	3	3	5	1	1
8	100.0	85.0	5	7	5	4	2
9	8.5.0	7.0	12	16			
10	16.0	11.2	12	15			
11	76.0	54.0	9	12			
12	189.0	134.0	9	12			
13	11.0	7.0	14	13	12	10	15
14	16.0	12.0	9	9	10	9	15
15	54.0	43.0	8	6	6	5	9
16	118.0	94.0	8	7	6	7	11
17	10.0	10.0	32	31			
18	30.0	55.0	23	30			
19	150.0	200.0	20	37			
20	195.0	188.0	25	27			
21	23.5.0	17.0	24	27	19	28	33
22	36.0	30.0	8	12	17	27	16
23	80.0	60.0	23	12	10	20	8
24	196.0	191.0	24	4	14	28	25

Tabla 4.2: Parámetros de los planes optimizados en la Tabla 4.1

No.	θ_0	θ_1	**a/b**	**N**	**ESC**	**ESC**$_{HW}$	α	α_{HW}	β	β_{HW}
1	0.935	0.64	10	2	2.2339	2.2339	0.0312	0.048	0.112	0.118
2	0.935	0.64	5	3	3.2136	3.2471	0.0475	0.048	0.117	0.118
3	0.935	0.64	1	6	10.008	10.321	0.0477	0.048	0.118	0.118
4	0.935	0.64	0.5	11	17.5384	18.2641	0.0480	0.048	0.118	0.118
5	0.935	0.64	10	3	2.1510	2.1852	0.0475	0.048	0.135	0.1353
6	0.935	0.64	5	3	2.9490	3.2113	0.0472	0.048	0.133	0.1353
7	0.935	0.64	1	5	9.4900	9.9041	0.0472	0.048	0.133	0.1353
8	0.935	0.64	0.5	8	16.61	17.5647	0.048	0.048	0.1353	0.1353
9	0.935	0.696	10	3	2.5405	2.6064	0.0550	0.0557	0.100	0.1
10	0.935	0.696	5	3	3.8901	3.9296	0.0547	0.0557	0.099	0.1
11	0.935	0.696	1	6	13.2820	13.4086	0.0537	0.0557	0.100	0.1
12	0.935	0.696	0.5	13	24.7597	24.7797	0.0557	0.0557	0.100	0.1
13	0.935	0.696	10	3	2.8822	2.9001	0.0371	0.0374	0.097	0.1
14	0.935	0.696	5	4	4.1527	4.1527	0.0369	0.0374	0.100	0.1
15	0.935	0.696	1	8	14.1412	14.3189	0.0370	0.0374	0.098	0.1
16	0.935	0.696	0.5	9	26.8122	26.8824	0.0356	0.0374	0.092	0.1
17	0.975	0.871	10	3	4.6820	4.682	0.0511	0.0519	0.099	0.1
18	0.975	0.871	5	3	7.8627	7.8989	0.0510	0.0519	0.099	0.1
19	0.975	0.871	1	8	29.835	30.6057	0.0518	0.0519	0.100	0.1
20	0.975	0.871	0.5	10	57.284	57.3572	0.0519	0.0519	0.100	0.1
21	0.975	0.871	10	3	4.756	4.7677	0.0402	0.0416	0.100	0.1
22	0.975	0.871	5	4	7.982	8.0274	0.0416	0.0416	0.100	0.1
23	0.975	0.871	1	7	31.817	34.4962	0.0415	0.0416	0.100	0.1
24	0.975	0.871	0.5	8	61.8304	62.6393	0.0416	0.0416	0.100	0.1

Tabla 4.3: Planes optimizados para la Tabla 2 en Hu and Wang (2018b)

No.	λ_0	λ_1	**Tamaños de grupos**												
1	8.98	8.07	10	7											
2	16.28	12.18	5	15	6										
3	62	40	4	2	4	4	3	3							
4	77	59	5	4	5	5	1	5	5	4	8	24	6		
5	8	5	5	12	4										
6	13	11	5	10	11										
7	48	37	5	10	11	37	12								
8	120	70	5	5	1	6	10	10	5	2					
9	9	11	12	8	7										
10	17	10	12	7	10										
11	46	46.5	8	5	8	4	7	9							
12	85.2	112.4	8	4	1	6	1	6	4	4	10	7	4	10	4
13	13	10	8	11	12										
14	9.96	11.41	9	8	8	8									
15	48	43	8	6	5	8	4	7	6	9					
16	123.75	88.5	8	2	4	2	5	6	6	4	9				
17	11.02	9.97	23	20	18										
18	25	24	20	20	18										
19	83.71	89.72	8	12	6	10	8	10	11	16					
20	195	205.33	8	12	2	5	8	8	19	15	9	17			
21	10	9.1	23	20	29										
22	29.25	27.70	20	23	5	23									
23	81.14	76.53	20	5	13	15	9	9	2						
24	226	258	8	10	1	16	2	18	20	17					

Tabla 4.4: Parámetros de los planes optimizados en la Tabla 4.3

	Propuesto					**Fleming**	
N	2	3	4	5	10	2	3
α	0.048	0.05	0.05	0.05	0.05	0.052	0.046
$1-\beta$	0.898	0.9	0.9	0.9	0.9	0.922	0.913
$\mathbf{ESS_0}$	26.0	25	24.7	24.7	24.0	32.5	31.6
$\mathbf{ESS_1}$	26.9	25.5	23.7	24.6	21.3	28	26.8
MSS	40	41	40	40	40	40	40
t	1.08	1.11	1.08	1.08	1.08	1.08	1.08
group	21	21	11	1	7	20	15
sizes	19	10	6	1	2	20	15
		10	5	15	1		10
			18	5	8		
				18	4		
					11		
					2		
					2		
					2		
					1		

Tabla 4.5: $\theta_0 = 0.05$ versus $\theta_1 = 0.2$, FSS= 37

	Propuesto					**Fleming**	
N	2	3	4	5	10	2	3
α	0.051	0.050	0.05	0.05	0.05	0.053	0.063
$1-\beta$	0.901	0.899	0.901	0.9	0.9	0.920	0.929
$\mathbf{ESS_0}$	25.7	22.1	20.8	19.3	20.3	24.7	24.8
$\mathbf{ESS_1}$	23.1	22.1	20.5	21.3	19.9	25.7	21.6
MSS	33	34	38	35	34	35	35
t	1.03	1.06	1.19	1.09	1.06	1.09	1.09
group	11	10	13	11	2	20	15
sizes	22	18	2	3	4	15	10
		6	4	12	1		10
			19	7	2		
				2	1		
					12		
					4		
					6		
					1		
					1		

Tabla 4.6: $\theta_0 = 0.1$ versus $\theta_1 = 0.3$, FSS= 32

	Propuesto					**Fleming**	
N	2	3	4	5	10	2	3
α	0.049	0.05	0.05	0.05	0.05	0.032	0.033
$1-\beta$	0.9	0.9	0.9	0.9	0.9	0.904	0.892
$\mathbf{ESS_0}$	31	30.9	31.2	28.4	29.2	39.3	33.8
$\mathbf{ESS_1}$	32.8	32.9	32.7	29.6	28.6	39.4	34.9
MSS	48	45	45	48	44	50	50
t	1.09	1.02	1.02	1.09	1.00	1.14	1.14
group	26	27	28	4	5	25	20
sizes	22	7	2	17	14	25	15
		11	11	15	7		15
			4	1	3		
				11	2		
					1		
					1		
					2		
					8		
					1		

Tabla 4.7: $\theta_0 = 0.2$ versus $\theta_1 = 0.4$, FSS= 44

	Propuesto					**Fleming**	
N	2	3	4	5	10	2	3
α	0.051	0.05	0.05	0.05	0.05	0.048	0.049
$1-\beta$	0.9	0.9	0.9	0.9	0.9	0.894	0.887
$\mathbf{ESS_0}$	35.8	32.9	32.0	32.2	28.4	37.1	31.6
$\mathbf{ESS_1}$	38.7	35.2	33.9	34.4	31.8	40.8	35.5
MSS	53	55	57	55	58	50	50
t	1.06	1.10	1.14	1.10	1.16	1.00	1.00
group	30	17	20	17	9	25	20
sizes	23	15	4	14	1	25	15
		23	13	1	3		15
			20	21	9		
				2	11		
					1		
					5		
					7		
					3		
					9		

Tabla 4.8: $\theta_0 = 0.3$ versus $\theta_1 = 0.5$, FSS= 50

N	Costo Fijo Adicional	Eficiencia (con optim.)	DE	Eficiencia (sin optim.)	DE	Efecto de la optimización
2	0	72.1	30.8	73.0	25.0	1.3
3		63.9	33.4	65.0	29.5	1.7
4		60.0	34.3	61.2	31.5	2.0
5		57.7	34.8	59.0	32.7	2.3
6		56.2	35.0	57.6	33.4	2.6
8		54.3	35.3	55.8	34.2	2.7
10		53.2	35.4	54.8	34.6	2.9
2	3	72.8	30.9	73.6	25.2	1.2
3		65.0	33.7	66.0	29.9	1.6
4		61.4	34.8	62.5	32.2	1.8
5		59.5	35.4	60.6	33.6	1.9
6		58.4	35.8	59.4	34.6	1.8
8		57.0	36.3	58.1	35.9	1.9
10		56.5	36.7	57.4	36.6	1.7
2	6	73.4	30.9	74.2	25.3	1.1
3		66.0	33.9	66.9	30.2	1.4
4		62.8	35.2	63.8	32.8	1.5
5		61.1	35.9	62.1	34.4	1.5
6		60.2	36.5	61.1	35.6	1.5
8		59.3	37.2	60.1	37.1	1.3
10		59.0	37.6	59.7	38.0	1.3
2	10	74.2	31.0	75.0	25.5	1.1
3		67.3	34.2	68.2	30.7	1.3
4		64.5	35.7	65.3	33.5	1.2
5		63.1	36.6	63.9	35.3	1.2
6		62.4	37.2	63.1	36.6	1.1
8		61.8	38.0	62.4	38.3	0.9
10		61.6	38.4	62.2	39.2	0.8
2	15	75.1	31.0	75.9	25.7	1.0
3		68.8	34.5	69.6	31.2	1.1
4		66.4	36.2	67.0	34.2	0.9
5		65.3	37.2	65.9	36.2	0.8
6		64.8	37.9	65.3	37.6	0.8
8		64.4	38.6	64.9	39.3	0.7
10		64.3	38.9	64.8	40.1	0.7

Tabla 4.9: Eficiencia de los planes óptimos en el marco bayesiano. $\alpha = \beta = 0.05$

N	Costo Fijo Adicional	Eficiencia (con optim.)	DE	Eficiencia (sin optim.)	DE	Efecto de la optimización
2	0	86.5	23.9	87.0	23.8	0.6
3		81.7	27.4	82.4	26.1	0.9
4		79.2	28.9	79.9	26.7	0.9
5		77.7	29.8	78.4	26.9	0.9
6		76.6	30.3	77.2	26.9	0.8
8		75.3	31.0	75.8	26.9	0.7
10		74.5	31.4	74.9	26.9	0.6
2	3	87.2	24.0	87.8	24.0	0.7
3		82.9	27.7	83.8	26.5	1.1
4		80.9	29.4	81.9	27.3	1.2
5		79.8	30.4	80.8	27.6	1.3
6		79.1	31.1	80.1	27.7	1.3
8		78.5	32.0	79.7	28.0	1.5
10		78.3	32.5	79.6	28.5	1.6
2	6	87.9	24.1	88.5	24.2	0.7
3		84.0	28.0	85.1	26.9	1.2
4		82.3	29.8	83.6	27.9	1.5
5		81.6	30.9	83.1	28.4	1.8
6		81.3	31.6	82.9	28.8	2.0
8		81.0	32.3	83.0	29.8	2.6
10		80.9	32.6	83.6	31.2	3.3
2	10	88.6	24.2	89.4	24.4	0.9
3		85.4	28.3	86.7	27.5	1.5
4		84.2	30.2	85.9	28.8	2.0
5		83.8	31.3	85.8	29.6	2.5
6		83.5	31.8	86.0	30.5	3.0
8		83.5	32.2	87.1	32.4	4.3
10		83.5	32.3	88.3	34.4	5.7
2	15	89.6	24.4	90.4	24.7	0.9
3		86.9	28.5	88.5	28.2	1.8
4		86.1	30.4	88.3	29.9	2.6
5		85.9	31.3	88.9	31.3	3.5
6		85.9	31.6	89.7	32.6	4.4

Tabla 4.10: Eficiencia de los planes óptimos para el problema de Kiefer-Weiss. $\alpha = \beta = 0.05$

Capítulo 5

Conclusiones y perspectivas

En esta última parte se presentan los resultados alcanzados durante la investigación, junto con propuestas y reflexiones que abren camino a estudios futuros.

En los capítulos 2–3 propusimos una solución computacional al problema de Kiefer-Weiss cuando se realiza el muestreo a partir de una familia exponencial de distribuciones (tanto en el caso continuo como en el discreto). Se desarrolló un conjunto de algoritmos para calcular los elementos de los planes de muestreo óptimos y sus características. Los algoritmos se implementaron como un código de programa en el lenguaje de programación R.

Se realizaron evaluaciones numéricas de los planes de muestreo óptimos y sus características, y se llevó a cabo una comparación numérica de los planes óptimos con la prueba SPRT y la prueba con tamaño de muestra fijo. Quedó claro que el patrón de eficiencia relativa de la prueba óptima de Kiefer-Weiss con respecto a la prueba de tamaño de muestra fijo y a la SPRT se mantiene en gran medida inalterado entre todos los modelos exponenciales que hemos investigado.

Hay un aspecto del problema de Kiefer-Weiss que dejamos sin tratar en los secciones anteriores: se trata del llamado 2-SPRT introducido por Lorden (1976). Este tiene un diseño simple basado en la ejecución simultánea de dos pruebas SPRT, y una definición natural de las regiones de continuación de cada SPRT hace que el desempeño general de la prueba sea óptimo asintóticamente.

El 2-SPRT se ha vuelto notable debido a su asombrosa eficiencia en el caso normal simétrico ($\alpha = \beta$) (ver Lorden, 1976). Pero ahora disponemos de herramientas bastante prácticas para el diseño óptimo, al menos para familias exponenciales, por lo que surgen preguntas sobre si la facilidad de uso del 2-SPRT justifica la pérdida de eficiencia en comparación con las soluciones verdaderamente óptimas de Kiefer-Weiss. Además, parece haber una falta de información sustancial sobre cómo se compara el desempeño del 2-SPRT con el de la solución óptima de Kiefer-Weiss en el caso general asimétrico (cuando $\alpha \neq \beta$), incluso para la distribución normal.

Para el caso general no simétrico, Huffman (1983) propuso una forma de definir θ para el 2-SPRT que lo hace eficiente asintóticamente, cuando $\alpha, \beta \to 0$ y $C_1 \leq \frac{\log \alpha}{\log \beta} \leq C_2$. Con los diseños óptimos que obtuvimos en este y en los secciones anteriores, es posible evaluar la eficiencia real (no asintótica) de los diseños 2-SPRT propuestos por Huffman (1983) y determinar el rango de valores de α, β y $\frac{\log \alpha}{\log \beta}$ para los cuales las fórmulas asintóticas de Huffman (1983) tienen una precisión aceptable en aplicaciones no asintóticas, para diferentes familias de distribuciones.

En el capítulo 4 presentamos un método para la construcción de pruebas secuenciales óptimas por grupos para observaciones que siguen una distribución perteneciente a una familia exponencial de un solo parámetro. Utilizamos como criterio de minimización el costo muestral esperado, ponderado entre un número finito de puntos del parámetro, suponiendo que el costo de un grupo de m observaciones es igual a $c_n(m)$ en la n-ésima etapa del procedimiento de prueba. Este enfoque incluye, como casos particulares, los marcos bayesiano y de Kiefer–Weiss, entre muchos otros.

El método propuesto está orientado al uso computacional, y hemos desarrollado un conjunto completo de algoritmos para calcular los planes óptimos y sus características de desempeño, así como una implementación informática de los algoritmos desarrollados en el lenguaje de programación R. El código del programa complementario abarca las distribuciones binomial y normal, y se encuentra disponible públicamente en un repositorio de GitHub en Internet (Farkhshatov and Novikov, 2025).

Para ilustrar la técnica propuesta, aplicamos el método en una serie de casos conocidos en la literatura. En particular, mostramos cómo los resultados de Hu and Wang (2018b) relativos a las formas de mejorar la eficiencia de una norma internacional para pruebas por atributos pueden complementarse mediante el uso de planes de prueba óptimos. Los resultados numéricos muestran que el uso de los planes óptimos propuestos, junto con una optimización adicional de los tamaños de grupo, permite mejorar la eficiencia de las pruebas en comparación con los planes casi óptimos de Hu and Wang (2018b).

En el contexto de los ensayos clínicos de Fase II (Fleming, 1982), aplicamos nuestros planes óptimos y comparamos numéricamente sus características de desempeño con las de los planes conocidos en la literatura. Observamos que la optimización adicional sobre el tamaño de grupo (de acuerdo con la Proposición 4.1.3) mejora drásticamente las características de desempeño (véanse Tablas 4.5 – 4.8).

Otra aplicación que consideramos fue en relación con el modelo de respuesta normal en el contexto de ensayos clínicos. Además de la evaluación de la eficiencia de los planes óptimos con respecto al plan no secuencial (de tamaño muestral fijo), analizamos el efecto de la optimización sobre el tamaño de los grupos y el efecto de la

variación de elementos de la estructura de costos. Los resultados numéricos (Tablas 4.9 y 4.10) muestran un desempeño tradicionalmente bueno de los planes secuenciales por grupos (con hasta 10 grupos), pero el efecto de la optimización del tamaño de grupo no es muy pronunciado. Para un número grande de grupos, recomendaríamos más bien el uso de planes planeados secuencialmente (pla adaptativos), que además evitan la necesidad de realizar una optimización multidimensional.

La técnica que utilizamos para la construcción principal en la sección 4.1.2 parece ser extensible a otros marcos secuenciales. En particular, es muy probable que las pruebas secuenciales óptimas por grupos para la distribución exponencial en el contexto de Hu and Wang (2018a) puedan obtenerse utilizando la misma técnica que en la sección 4.1.2. La construcción de pruebas planificadas secuencialmente óptimas también parece ser viable mediante una técnica similar.

Bibliografía

An, Q., Chang, M., 2016. Computation of optimal group sequential designs. Sequential Analysis 35, 453–464. URL: https://doi.org/10.1080/07474946.2016.1238256, doi:10.1080/07474946.2016.1238256, arXiv:https://doi.org/10.1080/07474946.2016.1238256.

Barber, S., Jennison, C., 2002. Optimal asymmetric one-sided group sequential tests. Biometrika 89, 49–60. URL: http://www.jstor.org/stable/4140558.

Bechhofer, R., 1960. A note on the limiting relative efficiency of the wald sequential probability ratio test. Journal of the American Statistical Association 55, 660—663. doi:10.1080/01621459.1960.10483366.

Bihorel, S., 2022. neldermead: R Port of the 'Scilab' Neldermead Module. URL: https://CRAN.R-project.org/package=neldermead. r package version 1.0-12.

Chang, M.N., 1996. Optimal designs for group sequential clinical trials. Communications in Statistics - Theory and Methods 25, 361–379. URL: https://doi.org/10.1080/03610929608831700, doi:10.1080/03610929608831700, arXiv:https://doi.org/10.1080/03610929608831700.

Chow, Y.S., Robbins, H., Siegmund, D., 1971. Great Expectations: The Theory of Optimal Stopping. Houghton Mifflin, Boston.

Dragalin, V., 2006. Adaptive Designs: Terminology and Classification. Drug Information Journal 40, 425–435. doi:10.1177/216847900604000408.

Eales, J.D., Jennison, C., 1992. An improved method for deriving optimal one-sided group sequential tests. Biometrika 79, 13–24. URL: http://www.jstor.org/stable/2337143.

Farkhshatov, F., Novikov, A., 2025. Group-sequential hypothesis testing: an R project for optimal test design. URL: https://github.com/tosinabase/groupsequential. accessed on 2025-05-14.

Fauß, M., Poor, H.V., 2020. Fading boundaries: On a nonparametric variant of the Kiefer–Weiss problem. arXiv:2010.12094.

Ferguson, T.S., 1967. Mathematical Statistics: A Decision Theoretic Approach. Academic Press, New York.

Fleming, T.R., 1982. One-Sample Multiple Testing Procedure for Phase II Clinical Trials. Biometrics 38.

Freeman, D., Weiss, L., 1964. Sampling plans which approximately minimize the maximum expected sample size. Journal of the American Statistical Association 59, 67–88.

Hawix, A., Schmitz, N., 1998. Remark on the modified Kiefer-Weiss problem for exponential families. Sequential Analysis 17, 297—303.

Hu, S., Wang, H., 2018a. A heuristic approach for near optimal truncated sequential test of exponential distribution. Sequential Analysis 37, 431–454. URL: https://doi.org/10.1080/07474946.2018.1554887.

Hu, S., Wang, H., 2018b. Nearly optimal truncated group sequential test on binomial proportions. Communications in Statistics - Simulation and Computation 47, 2332–2342. URL: https://doi.org/10.1080/03610918.2017.1343837, doi:10.1080/03610918.2017.1343837, arXiv:https://doi.org/10.1080/03610918.2017.1343837.

Huffman, M.D., 1983. An efficient approximate solution to the Kiefer-Weiss problem. The Annals of Statistics 11, 306–316.

ISO2859-1, 1999. ISO 2859-1. Sampling Procedures for Inspection by Attribute - Part 1. International Organization for Standardization.

Jennison, C., 1987. Efficient group sequential tests with unpredictable group sizes. Biometrika 74, 155–165. URL: http://www.jstor.org/stable/2336030.

Jennison, C., Turnbull, B.W., 1999. Group Sequential Methods with Applications to Clinical Trials. Chapman and Hall/CRC.

Kiefer, J., Weiss, L., 1957. Some properties of generalized sequential probability ratio tests. Annals of Mathematical Statistics 28, 57–75.

Lai, T.L., 1973. Optimal stopping and sequential tests which minimize the maximum expected sample size. The Annals of Statistics 1, 659–673.

Lorden, G., 1976. 2-SPRT'S and The Modified Kiefer-Weiss Problem of Minimizing an Expected Sample Size. The Annals of Statistics 4, 281 – 291. URL: https://doi.org/10.1214/aos/1176343407, doi:10.1214/aos/1176343407.

Lorden, G., 1980. Structure of sequential tests minimizing an expected sample size. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 51, 291–302.

Murdoch, D., Adler, D., 2021. rgl: 3D Visualization Using OpenGL. Https://github.com/dmurdoch/rgl, https://dmurdoch.github.io/rgl/.

Nelder, J.A., Mead, T., 1965. A Simplex Method for Function Minimization. Computer Journal 7, 308–313. doi:doi:10.1093/comjnl/7.4.308.

Novikov, A., 2009a. Optimal sequential tests for two simple hypotheses. Sequential Analysis 28, 188–217.

Novikov, A., 2009b. Optimal sequential tests for two simple hypothesis. Sequential Analysis 28, 188–217.

Novikov, A., 2024a. Group sequential hypothesis tests with variable group sizes: optimal design and performance evaluation. Comm. Statist. Theory Methods 53, 5744–5760. URL: https://doi.org/10.1080/03610926.2023.2231155, doi:10.1080/03610926.2023.2231155.

Novikov, A., 2024b. Gsmultihypothesis: an R project for computing optimal group sequential multi-hypothesis tests for exponential families of distributions. URL: https://github.com/HOBuKOB-MEX/GSmultihypothesis. gitHub repository. Accessed 2025-05-14.

Novikov, A., Novikov, A., Farkhshatov, F., 2021. An R project for numerical solution of the Kiefer-Weiss problem. https://github.com/tosinabase/Kiefer-Weiss. https://github.com/tosinabase/Kiefer-Weiss.

Novikov, A., Novikov, A., Farkhshatov, F., 2022. A computational approach to the kiefer-weiss problem for sampling from a bernoulli population. Sequential Analysis 41, 198–219. URL: http://www.tandfonline.com/doi/abs/10.1080/07474946.2022.2070212, doi:10.1080/07474946.2022.2070212.

Novikov, A., Novikov, A., Farkhshatov, F., 2023. Numerical solution of kiefer-weiss problems when sampling from continuous exponential families. Sequential Analysis 42, 189–209. URL: https://doi.org/10.1080/07474946.2023.2193602, doi:10.1080/07474946.2023.2193602.

Novikov, A., Popoca Jiménez, X., 2022. Optimal group-sequential tests with groups of random size. Sequential Analysis 41, 220–240. URL: http://www.tandfonline.com/doi/abs/10.1080/07474946.2022.2070213, doi:10.1080/07474946.2022.2070213.

Pocock, S.J., 1977. Group sequential methods in the design and analysis of clinical trials. Biometrika 64, 191–199. URL: https://doi.org/10.1093/biomet/64.2.191, doi:10.1093/biomet/64.2.191, arXiv:https://academic.oup.com/biomet/article-pdf/64/2/191/1089682/64-2-191.pdf.

Schmitz, N., 1993. Optimal Sequentially Planned Decision Procedures. Lecture Notes in Statistics 79, Springer-Verlag, New York.

Sekhon, J.S., Mebane, W. R. Jr., 2011. Genetic optimization using derivatives: The rgenoud package for R. Journal of Statistical Software 42, 1–26. URL: https://www.jstatsoft.org/v42/i11/.

Tarima, S., Flournoy, N., 2022. Most Powerful Test Sequences with Early Stopping Options . Metrika 85, 491–513. URL: https://doi.org/10.1007/s00184-021-00839-w.

Tartakovsky, A.G., Nikiforov, I.V., Basseville, M., 2014. Sequential analysis: hypothesis testing and changepoint detection. Chapman & Hall/CRC Press, Boca Raton, Florida.

Wald, A., 1945. Sequential tests of statistical hypotheses. The Annals of Mathematical Statistics 16, 117–186. URL: http://www.jstor.org/stable/2235829.

Wald, A., Wolfowitz, J., 1948. Optimum character of the sequential probability ratio test. Annals of Mathematical Statistics 19, 326–339.

Weiss, L., 1962. On sequential tests which minimize the maximum expected sample size. Journal of American Statistical Assocciation 57, 551–566.

Young, L.J., 1994. Computation of some exact properties of Wald's SPRT when sampling from a class of discrete distributions. Biom. J. 36, 627–637.

Casa abierta al tiempo
UNIVERSIDAD AUTÓNOMA METROPOLITANA

ACTA DE DISERTACIÓN PÚBLICA

No. 00099
Matrícula: 2212801602

Pruebas de hipótesis secuenciales: aspectos computacionales de optimización y aplicaciones.

En la Ciudad de México, se presentaron a las 12:00 horas del día 16 del mes de diciembre del año 2025 en la Unidad Iztapalapa de la Universidad Autónoma Metropolitana, los suscritos miembros del jurado:

DR. JOSE ANDRES CHRISTEN GRACIA
DR. EVGUENI GORDIENKO
DR. ANDREY NOVIKOV
DR. HUGO ADAN CRUZ SUAREZ
DR. ALEJANDRO ROMAN VASQUEZ

Bajo la Presidencia del primero y con carácter de Secretario el último, se reunieron a la presentación de la Disertación Pública cuya denominación aparece al margen, para la obtención del grado de:

DOCTOR EN CIENCIAS (MATEMÁTICAS)

DE: FAIL FARKHSHATOV

y de acuerdo con el artículo 78 fracción IV del Reglamento de Estudios Superiores de la Universidad Autónoma Metropolitana, los miembros del jurado resolvieron:

Aprobar

Acto continuo, el presidente del jurado comunicó al interesado el resultado de la evaluación y, en caso aprobatorio, le fue tomada la protesta.



FAIL FARKHSHATOV
ALUMNO

REVISÓ

MTRA. ROSALÍA SERRANO DE LA PAZ
DIRECTORA DE SISTEMAS ESCOLARES

DIRECTOR DE LA DIVISIÓN DE CBI	PRESIDENTE	VOCAL
DR. ROMAN LINARES ROMERO	DR. JOSE ANDRES CHRISTEN GRACIA	DR. EVGUENI GORDIENKO

VOCAL	VOCAL	SECRETARIO
DR. ANDREY NOVIKOV	DR. HUGO ADAN CRUZ SUAREZ	DR. ALEJANDRO ROMAN VASQUEZ