



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA

ACTA DE DISERTACIÓN PÚBLICA

No. 00080

Matrícula: 2142800700

Pruebas de Hipótesis
Secuenciales bajo Horizonte
Aleatorio.

En la Ciudad de México, se presentaron a las 13:00 horas del día 14 del mes de octubre del año 2022 en la Unidad Iztapalapa de la Universidad Autónoma Metropolitana, los suscritos miembros del jurado:

DR. JOSE ANDRES CHRISTEN GRACIA
DR. HUGO ADAN CRUZ SUAREZ
DR. ANDREY NOVIKOV
DR. EVGUENI ILICH GORDIENKO
DR. JOSE RAUL MONTES DE OCA MACHORRO

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 (MATEMATICAS)

DE: JUAN LUIS PALACIOS SOTO

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.

REVISÓ

MTRA. ROSALÍA SERRANO DE LA PAZ
DIRECTORA DE SISTEMAS ESCOLARES

DIRECTOR DE LA DIVISIÓN DE CBI

Roman Linares Romero
DR. ROMAN LINARES ROMERO

PRESIDENTE

Jose Andres Christen Gracia
DR. JOSE ANDRES CHRISTEN GRACIA

VOCAL

Hugo Adan Cruz Suarez
DR. HUGO ADAN CRUZ SUAREZ

VOCAL

Andrey Novikov
DR. ANDREY NOVIKOV

VOCAL

Evgueni Ilich Gordienko
DR. EVGUENI ILICH GORDIENKO

SECRETARIO

Jose Raul Montes de Oca Machorro
DR. JOSE RAUL MONTES DE OCA
MACHORRO



UNIVERSIDAD AUTÓNOMA METROPOLITANA
IZTAPALAPA

DIVISIÓN DE CIENCIAS BÁSICAS E INGENIERÍA

Pruebas de Hipótesis Secuenciales bajo Horizonte Aleatorio

TESIS

QUE PARA OBTENER EL TÍTULO DE
DOCTOR EN CIENCIAS MATEMÁTICAS

PRESENTA

JUAN LUIS PALACIOS SOTO

MATRÍCULA: 2142800700

ASESOR

ANDREY NOVIKOV

Resumen

Las pruebas secuenciales se han aplicado, bajo el debido rigor, desde la Segunda Guerra Mundial, con la investigación de [Wald \(1947\)](#). Hoy en día su aplicación en la industria es muy variada y representa un ahorro en la inferencia sin perder la calidad en la misma. Sin embargo, la aparición de un evento externo al experimento secuencial puede alterar el rumbo natural de una prueba secuencial sin haber tomado la decisión final, lo que hace el análisis interesante, tanto teórico como aplicado. En esta investigación hemos aplicado la teoría clásica a las pruebas secuenciales bajo un horizonte aleatorio con distribución geométrica y hemos obtenido resultados más generales; además de que hemos simulado de manera computacional algunos resultados.

DEDICATORIA

Por toda tu paciencia y amor en todos estos años Yanet
Y desde luego a nuestros motores: Iyari y Samuel.

Agradecimientos

Primeramente deseo agradecer profundamente al Dr. Andrey Novikov, por su apoyo y asesoría para la realización de este trabajo. Gracias por compartirme parte de su conocimiento. Extiendo mi agradecimiento a todos mis sinodales: el Dr. Raúl Montes de Oca, Dr. Hugo Adán Cruz, Dr. Gordienko y al Dr. Christen por el tiempo que se tomaron en revisar mi trabajo y sus valiosas observaciones. Deseo agradecer al CONACYT que sin su apoyo económico durante mis estudios de posgrado en la UAM-I, éstos no hubieran sido posible. No puedo dejar fuera de mis agradecimientos a la UAM que me ofreció varios y buenos años de aprendizaje, además de que ahí conocí a mi esposa amada Yanet. Finalmente, agradezco los ratos que mi amigo Adrián Vázquez me molestaba con terminar la tesis a pesar de que mis ánimos eran nulos para continuar con esta investigación.

Índice general

Resumen	I
Agradecimientos	v
Introducción	ix
1. Teoría Secuencial	1
1.1. Antecedentes	1
1.2. Conceptos básicos	8
1.3. Planteamiento del problema	10
2. Pruebas Óptimas bajo Horizonte	15
2.1. Reducción a un problema de paro óptimo	15
2.2. Reglas de paro óptimas	19
2.2.1. Reglas de Paro Óptimo Truncadas	19
2.2.2. Optimalidad de las reglas de paro no truncadas	21
3. Optimalidad para Observaciones Independientes	29
3.1. El problema Kiefer-Weiss bajo horizonte aleatorio	29
3.2. Minimización del promedio muestral bajo una hipótesis	35
3.3. Observaciones i.i.d	41
4. Aplicaciones	49
4.1. Pruebas secuenciales de Bayes para dos hipótesis compuestas	49
4.2. Procesos de Markov a tiempo discreto	52
4.3. Simulaciones Computacionales	55
4.3.1. Estimando: $\alpha(\psi, \phi)$, $\beta(\psi, \phi)$, $E_0(H \wedge n)$ y $E_1(H \wedge n)$	55
4.3.2. Obtención de la Región de Continuación	62
5. Conclusiones	77
Apéndice A. Demostraciones	79
A.1. Demostración del Teorema 2.1	79

	79
A.2. Demostración del Lema 2.3	79
A.3. Demostración del Lema 2.4	81
A.4. Demostración del Corolario 2.3	81
A.5. Demostración del Lema 2.5	82
A.6. Demostración del Lema 3.1	83
A.7. Demostración del Lema 3.2	83
A.8. Demostración del Lema 3.3	83
A.9. Demostración del Lema 3.6	85
A.10. Demostración del Teorema 3.4	86
Apéndice B. Programas en GNU Octave	91
B.1. Código para estimar errores tipo I, II y $E_i(\min\{H, n\})$	91
B.2. Código de la función para determinar la región de continuación (A_n, B_n) bajo horizonte H	95

Introducción

Dentro de la inferencia estadística se encuentra un método desarrollado durante la Segunda Guerra Mundial que se denomina *análisis secuencial* y una de sus aplicaciones es a las pruebas de hipótesis. En las pruebas de hipótesis que hoy en día conocemos, uno de los problemas a considerar tiene que ver con el tamaño n de la muestra para realizar dicha prueba, debido a que si es muy pequeña puede ser insuficiente la información como para tomar una decisión; mientras que si es demasiado grande, nos puede ocasionar un gasto excesivo. Se sabe que si fijamos las probabilidades de error tipo I y II es posible determinar, conociendo desde luego la distribución del o los estadísticos involucrados, el tamaño mínimo N sugerido para realizar la prueba con un nivel de confianza del $(1 - \alpha)$, donde α es la cota de la probabilidad de cometer el error tipo I. En una prueba de hipótesis secuencial, la característica principal es que el tamaño de la muestra no se fija de antemano, sino que es vista como una variable aleatoria, la cual depende de las observaciones tomadas con anterioridad. Un procedimiento secuencial es de suma importancia en muchos estudios aplicados ya que representa un ahorro en el gasto del experimento porque nos lleva a una decisión sobre alguna de las hipótesis planteadas con el mínimo de observaciones, alrededor del 50 % ($N/2$) con respecto a las pruebas de tamaño fijo y son además igual de confiables (Wald, 1947).

Un método en particular de este análisis secuencial es conocido hoy en día como la *Prueba Secuencial en Razón de Probabilidades* (SPRT por sus siglas en inglés), el cual fue presentado con detalle en Wald (1947). Sin embargo, la optimalidad de la SPRT se presentó un año después en Wald y Wolfowitz (1948). Hoy en día sabemos que tanto la teoría secuencial como en particular la SPRT tienen variadas aplicaciones dentro de la industria, neurociencia, medicina, por mencionar algunos ejemplos (Bartroff et al., 2013; Closas et al., 2012; Dayanik y Yu, 2013; Frazier y Yu, 2008; Wald, 1947). Sabemos que en general una prueba secuencial como la SPRT, consta de dos reglas: una llamada regla de paro y otra llamada regla de decisión. Es así que nos surge la inquietud de analizar qué sucedería si un experimento secuencial se ve afectado por un factor externo a las mismas observaciones y no por la regla de paro de la prueba secuencial definida para un experimento, afectando la decisión final. Desde luego que este planteamiento tiene sentido práctico. Por ejemplo, podemos suponer que cierto experimento o proceso se ve afectado por cuestiones de presupuesto, teniendo que reportar cuál será la mejor decisión con la información que se tenga en ese momento. Entonces tiene sentido práctico suponer que el horizonte que gobierna al proceso es aleatorio e independiente de las observaciones del proceso, ya que la toma de observaciones puede ser interrumpida por falta de recursos, porque el tiempo para la decisión final se ha terminado o simplemente por un evento ajeno

a nuestra voluntad y externo a las observaciones. Por supuesto que en la teoría de pruebas secuenciales, se cuenta con variada información con este enfoque del horizonte aleatorio, pero lo que nosotros planteamos en esta investigación es completamente original, a diferencia por ejemplo de [Dayanik y Yu \(2013\)](#) y [Frazier y Yu \(2008\)](#) en donde ellos después de que se interrumpe el experimento ya no se hace la decisión final y en cambio se amonesta porque no se llegó a una decisión final, pero es entendible porque el enfoque que le dan a las pruebas son aplicadas a la neurociencia.

La forma en que está estructurada la siguiente investigación es como sigue: en el Capítulo 1, explicaremos brevemente los antecedentes que consideramos más relevantes, siguiendo con algunos conceptos básicos tanto de la teoría clásica secuencial como no secuencial como vemos al revisar el Lema de Neyman-Pearson; asimismo recordaremos en qué consiste la SPRT y las generalizaciones de ésta que han realizado algunos investigadores. También expondremos algunos detalles técnicos referentes a procesos estocásticos, como son filtraciones y tiempos de paro. Finalmente, realizaremos el planteamiento principal de la presente investigación bajo el enfoque del horizonte aleatorio.

Para el Capítulo 2, resolveremos por medio de programación dinámica el problema principal, que hemos llamado *Kiefer-Weiss Modificado Bajo Horizonte Aleatorio* (KWMH), primero para el llamado caso truncado, y posteriormente extenderlo al caso no truncado. Caracterizaremos completamente la optimalidad de las pruebas secuenciales bajo el horizonte aleatorio. Cabe mencionar que a partir del Capítulo 2 hasta el Capítulo 4, los resultados son nuestra contribución al análisis secuencial, a excepción de los lemas [2.1](#) y [2.2](#). Un ejemplo que ilustra esta teoría lo pueden ver en la Sección [3.1](#).

En el Capítulo 3, aplicaremos los resultados del Capítulo 2 para cuando las observaciones son únicamente independientes; cuando las observaciones son i.i.d, pero suponiendo que el horizonte aleatorio posee distribución geométrica, obteniendo resultados importantes relacionados con la teoría clásica secuencial. Finalmente abordaremos el que hemos denominado el problema inverso.

En el Capítulo 4 mostramos algunos resultados aplicados a las pruebas secuenciales del tipo Bayes para dos pruebas de hipótesis compuestas. También aplicamos nuestros resultados para caracterizar procesos de Markov a tiempo discreto. Y finalmente, hemos realizado simulación numérica con el Método Montecarlo para observaciones discretas y continuas: del tipo Bernoulli y exponencial. Principalmente hemos hecho dos tipos de simulaciones computacionales. Primero, para estimar los errores tipo I y II cuando fijamos el intervalo (A, B) , conocido como región de continuación; asimismo para estimar el tiempo promedio de paro del experimento, cuando generamos observaciones, discretas o continuas, bajo cada una de las hipótesis, en donde la tasa de fallos debido al horizonte aleatorio es nuestro dominio. Segundo, y conocido como el problema inverso, para determinar la región de continuación cuando suponemos que el horizonte tiene distribución geométrica en una prueba de hipótesis simple cuyas observaciones son del tipo Bernoulli, variando la tasa de fallos p_H debido al horizonte aleatorio. En esta parte quedan muchas interrogantes que se pueden tratar en un trabajo futuro, como son los parámetros: de costo $c > 0$ y el valor K

para decidir por una de las hipótesis cuando el experimento termina debido al horizonte aleatorio y no por la regla de paro.

Capítulo 1

Teoría Secuencial

1.1. Antecedentes

En este primer capítulo, hablaremos de la parte motivacional que dio origen a este trabajo, y para ello, revisaremos brevemente algunos antecedentes de la teoría clásica, tanto secuencial como no secuencial y algunas variantes a lo largo de su historia.

Un primer acercamiento a las pruebas secuenciales se presentó con [Dodge y Romig \(1929\)](#) quienes construyeron un procedimiento de muestreo doble en el cual el número de observaciones no se determina de antemano, pero que depende de las mismas observaciones. Este método de doble muestreo, consiste básicamente en tomar una primera muestra de tamaño n_1 y de acuerdo a esta primera muestra se decide si se toma una segunda muestra o no de tamaño n_2 . Posteriormente [Bartky \(1943\)](#) diseñó un muestreo múltiple pero para algunos casos en concreto.

Una motivación que dio origen a la *Prueba Secuencial en Razón de Probabilidades* (SPRT por sus siglas en inglés) de Wald, fue el Lema de Neyman-Pearson. Estos personajes, Neyman y Pearson, establecieron y demostraron un método para construir lo que se conoce hoy en día como la prueba más potente para una prueba simple frente a otra prueba simple también. Supongamos que X tiene una función de densidad $f(x; \theta)$, o función de probabilidad para el caso discreto, y deseamos probar $H_0 : \theta = \theta_0$ contra $H_1 : \theta = \theta_1$, $\theta_0 < \theta_1$ ($\Theta = \{\theta_0\} \cup \{\theta_1\}$ es un espacio paramétrico que consta de únicamente dos elementos). El lema se establece como sigue:

Lema 1.1 (Neyman and Pearson, 1933). Sean $X_1, X_2, \dots, X_n$ una muestra aleatoria y sea

$$l_n = \frac{\prod_{i=1}^n f(X_i, \theta_1)}{\prod_{i=1}^n f(X_i, \theta_0)}. \quad (1.1)$$

Entonces la prueba más potente de H_0 en contra de H_1 se obtiene al rechazar H_0 si $l_n \geq K$, y se acepta H_0 si $l_n < K$, donde K se determina por el nivel de significancia.

La prueba más potente radica básicamente en fijar el error tipo I y minimizar el error tipo II suponiendo de antemano una muestra de tamaño n fijo. En esto radica la diferencia con respecto a las pruebas secuenciales, el de no fijar el tamaño de la muestra, y lo cual hace a este tipo de pruebas óptimas al utilizar menos recursos ([Wald y Wolfowitz, 1948](#)).

La llamada SPRT del análisis secuencial clásico consiste en lo siguiente. Supongamos que estamos tomando observaciones $X_1, X_2, X_3, \dots$ i.i.d. y que provienen de una función de probabilidad (de masa para el caso discreto y densidad para el caso continuo) en común $f(\cdot; \theta)$. De ahora en adelante sólo le llamaremos función de probabilidad, donde θ es un parámetro. Podemos suponer que θ es un sólo parámetro, sin embargo, en la teoría secuencial se aborda el problema general donde θ es visto como un vector de parámetros. El problema principal consiste en probar

$$H_0 : \theta = \theta_0 \quad \text{contra} \quad H_1 : \theta = \theta_1, \quad \theta_0 < \theta_1,$$

minimizando

$$E_i(N), i = 0, 1,$$

donde N corresponde al tamaño máximo de la muestra $\{X_1, \dots, X_N\}$.

Sujeto a

$$P(\text{rechazar } H_0 | \theta_0) \leq \alpha \quad \text{y} \quad P(\text{aceptar } H_0 | \theta_1) \leq \beta,$$

donde $\alpha, \beta \in (0, 1)$ son las cotas para los errores tipo I y tipo II, respectivamente. Ahora bien, para aplicar la SPRT se eligen dos números A, B cualesquiera, tales que cumplan $0 < A \leq 1 \leq B < \infty$ y $A \neq B$. Sea l_n definido como en (1.1), entonces para $n = 1$, tenemos que ver si se verifica lo siguiente:

1. si $l_n \leq A$ aceptamos H_0 ,
2. si $B \geq l_n$ aceptamos H_1 ,
3. si $A < l_n < B$ tomamos una nueva observación y repetimos lo anterior con l_{n+1} .

El intervalo (A, B) se conoce como el **intervalo o región de continuación** (continuar observando datos o con el muestreo). Una pregunta que surge de manera natural es, ¿qué tan arbitrarios deben ser los valores A y B que se eligen para iniciar la prueba SPRT? Pues bien, Wald trabajó en este tema y encontró unas cotas para dichas constantes, que se conocen hoy en día como cotas de Wald, y que en la mayoría de los propósitos prácticos son suficiente (Wald, 1947). Así tenemos que si α y β son las cotas de nuestros errores tipo I y II, respectivamente, entonces los valores A y B para la SPRT tienen las siguientes cotas:

$$\frac{\beta}{1 - \alpha} \leq A < B \leq \frac{1 - \beta}{\alpha}.$$

Una justificación por utilizar un método secuencial, como ya mencionamos, es debido a que minimiza el tamaño muestral esperado con respecto a la prueba más potente. De esta manera tenemos un procedimiento secuencial que optimiza los recursos. Se probó que esta minimización del tamaño esperado se satisface bajo ambas hipótesis y que además, un procedimiento de este tipo es acotado y tiene terminación finita con probabilidad 1 (Wald y Wolfowitz, 1948).

La optimalidad de la SPRT se presentó en 1948 bajo el supuesto de que $X_1, X_2, \dots$ es una sucesión de variables aleatorias i.i.d. cuya función de probabilidad está dada por $f(x; \theta)$.

Teorema 1.1 (Wald and Wolfowitz, 1948). *De entre todas las pruebas (de tamaño fijo o secuencial) para las cuales $P(\text{rechazar } H_0|\theta_0) < \alpha$, $P(\text{aceptar } H_0|\theta_1) < \beta$ y tales que $E(N|\theta_i) < \infty$, $i = 0, 1$, la SPRT con probabilidad de errores α y β , minimiza tanto a $E(N|\theta_0)$ como a $E(N|\theta_1)$.*

Para demostrar este hecho utilizaron un enfoque bayesiano, donde minimizan lo que se conoce como riesgo bayesiano de un procedimiento δ , dado por

$$r(\pi, w_0, w_1, \delta) = \pi[w_0\alpha + cE(N|\theta_0)] + (1 - \pi)[w_1\alpha + cE(N|\theta_1)], \quad (1.2)$$

donde θ es una variable aleatoria tal que $P(\theta = \theta_0) = \pi = 1 - P(\theta = \theta_1)$; c representa el costo por observación y w_i $i = 0, 1$ son los pesos que se le asigna a las probabilidades de los errores. Para demostrar el teorema anterior se hace mención de un cierto procedimiento δ , que dentro de la teoría secuencial se conoce hoy en día como prueba secuencial.

Algo que debemos tener en cuenta en las pruebas de hipótesis es la forma de ver el planteamiento de una de éstas. Por ejemplo para pruebas paramétricas, es muy común hacer pruebas de hipótesis acerca de los parámetros, conocida la distribución. Digamos que si $X_1, X_2, \dots, X_n$ es una muestra aleatoria secuencial tomada al tiempo n y cuya densidad conjunta verdadera está dada por $f(x_1, \dots, x_n; \theta)$ (en el caso continuo o función de probabilidad en el caso discreto), donde θ es el parámetro verdadero desconocido que, en el caso de pruebas de hipótesis simples, puede ser θ_0 o bien θ_1 . Así, si se desea hacer estimación sobre el parámetro θ de esta distribución, entonces un planteamiento de una prueba de hipótesis simple, puede ser planteada como sigue:

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta = \theta_1.$$

Sin embargo uno podría plantear lo siguiente:

$$H_0 : f = f_0 \quad \text{vs} \quad H_1 : f = f_1,$$

donde f_i es igual a la densidad conjunta bajo la hipótesis $i = 0, 1$. Incluso podríamos plantear

$$H_0 : P = P_0 \quad \text{vs} \quad H_1 : P = P_1,$$

donde P es la distribución verdadera de las observaciones, mientras que P_i es la distribución de las observaciones bajo las hipótesis $i = 0, 1$.

Después de la aparición de la optimización de la SPRT hubo modificaciones, tanto de los mismos autores como de otros interesados en el problema. Por ejemplo, Weiss (1953) mostró una generalización de la SPRT (GSPRT, por sus siglas en inglés), donde la diferencia principal radica en que el intervalo de continuación (A, B) no es el mismo durante el experimento, sino que las constantes son diferentes en cada paso, es decir, (A_n, B_n) , con $A_n < B_n$ para el paso n . Además, supone que las densidades $f(x_i, \theta_j)$, $j = 0, 1$, bajo las hipótesis H_0 y H_1 son acotadas casi en todas partes, es decir, excepto en un conjunto de probabilidad nula. También abordó el caso truncado, esto es, que existe un horizonte finito de observaciones, digamos N .

En el trabajo de [Kiefer y Weiss \(1957\)](#) se trabaja con un aspecto muy importante con respecto al número muestral esperado que se calcula suponiendo que el flujo de información proviene de una distribución diferente de las hipotetizadas. Por ejemplo, si se desea minimizar el valor esperado muestral, $E_\theta(N)$, bajo la prueba

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta = \theta_1,$$

con errores tipo I y II acotados por α y β , respectivamente, entonces se minimiza bajo una tercera distribución, cuyo parámetro verdadero es θ y que no necesariamente coincide con los parámetros hipotetizados. Este problema es bien conocido con el nombre del *problema Kiefer-Weiss modificado*.

Otro trabajo, no menos importante lo vemos en [Weiss \(1962\)](#), en donde se interesó en demostrar que dada una sucesión de variables aleatorias $X_1, X_2, X_3, \dots$ con distribución común, excepto quizás para algún parámetro θ , el problema consistió en probar una hipótesis simple $H_{-1} : \theta = \theta_{-1}$ contra un hipótesis alternativa simple también $H_1 : \theta = \theta_1$, donde $\theta_{-1} < \theta_1$, y restringiéndose a pruebas con errores tipo I y II acotados por dos valores α y β ($\alpha + \beta < 1$) dados. La principal característica en este trabajo es que se minimizó el tamaño muestral esperado, $E(N|\theta)$, bajo un tercer parámetro θ . También se minimizó el máximo tamaño muestral esperado, sobre todas las pruebas T que cumplen ciertas condiciones, es decir, $\min_T \max_\theta E(N|\theta)$. Se trabajó con dos ejemplos en particular y en ambos se supone el caso simétrico que consiste simplemente en fijar $\alpha = \beta$; pero en el primero se analizan observaciones provenientes de una distribución normal con media θ desconocida y varianza σ^2 conocida, $N(\theta, \sigma^2)$; mientras que en el caso 2, se suponen observaciones con distribución común de Bernoulli, tal que $P(x = 0) = 1 - \theta$ y $P(x = 1) = \theta$. Para lograr su objetivo construyó una prueba bayesiana T con respecto a una distribución a priori (b_{-1}, b_0, b_1) , tal que $b_{-1} + b_0 + b_1 = 1$ y donde las probabilidades de error tipo I y II para T están acotadas por dos valores α y β . Entonces minimiza $E(N|\theta)$ bajo una tercera distribución conocida con parámetro θ_0 tal que $\theta_{-1} < \theta_0 < \theta_1$ como en [Kiefer y Weiss \(1957\)](#). De hecho la cantidad

$$b_{-1}P_{\theta_{-1}}[H_{-1} \text{ sea rechazado}] + b_0E(N|\theta_0) + b_1P_{\theta_1}[H_1 \text{ sea rechazado}],$$

es minimizada por T .

Ahora bien, para minimizar el máximo tamaño muestral esperado, se parte del supuesto de que $E(N|\theta \text{ usando } T) = \max_\theta E(N|\theta \text{ usando } T)$ y se procede por contradicción en su demostración. En dicho trabajo se analiza también el caso truncado, es decir, hay un máximo de observaciones por tomar en el experimento y en caso de no terminar el experimento, ya sea por la prueba o regla en cuestión, se decide por una de las dos pruebas, básicamente eligiendo a placer, 50-50.

En el trabajo de [Burkholder y Wijsman \(1963\)](#) se menciona que para demostrar la propiedad de optimalidad en la SPRT dada por [Wald y Wolfowitz \(1948\)](#), se hace bajo dos condiciones: una que los valores A, B satisfacen $A \leq 1 \leq B$, y dos, que sólo las pruebas secuenciales con tamaño muestral finito se tomaron en cuenta. En este trabajo demuestran que la segunda condición no es necesaria, no así la primera a menos que se consideren pruebas con al menos una observación. También se trabaja bajo un costo por

observación, el cual no se fija de antemano, lo cual en la práctica tiene sentido, debido a que los costos pueden variar con el tiempo.

Otra variante que apareció fue el trabajar con pruebas secuenciales aleatorizadas, (DeGroot, 1970; Ferguson, 1967). En estos trabajos se analizan las propiedades de las pruebas de decisión (secuenciales). Se explican ciertas ventajas de hacer uso de estas reglas aleatorizadas, como por ejemplo tiene la propiedad de linealidad, insesgadez, suficiencia, etcétera.

Se presentó una nueva generalización en Eisenberg et al. (1976), al considerar observaciones no necesariamente independientes e idénticamente distribuidas para las pruebas del tipo GSPRT's. Se centran en el estudio de una amplia gama de pruebas que contienen a las pruebas SPRT's. Describen que el modelo basado en observaciones i.i.d. es inapropiado toda vez que las razones de verosimilitud que se utilizan en las pruebas secuenciales no están definidas directamente de las muestras aleatorias. Para ello, definen una secuencia anidada, que no es mas que una filtración, dada de manera natural por las mismas observaciones y por la variable de paro N . En este nuevo planteamiento se tiene como caso particular el Lema de Neyman-Pearson.

En Cochlar, J., Vrana, I (1978) se toman observaciones de tamaño M en cada etapa n ($X_n^M = (X_{n1}, X_{n2}, \dots, X_{nM})$, al tiempo o etapa n) y en donde suponen que los vectores de observaciones de etapa en etapa no tienen por que ser idénticamente distribuidos ($X_1^M, X_2^M, X_3^M, \dots$). Sin embargo, para una sola etapa n , las observaciones $X_{n1}, X_{n2}, \dots, X_{nM}$ sí se suponen idénticamente distribuidas, aunque no independientes, pero sí independientes de un vector a otro. Se aborda este problema tomando la clase GSPRT, por lo que se tienen cotas distintas en cada etapa.

En el trabajo de Lorden (1980), se trabaja básicamente con las condiciones clásicas en la teoría secuencial, es decir, se pretende minimizar el tamaño muestral esperado, $E(N)$, en la prueba $H_0 : p = f$ vs $H_1 : p = g$, con observaciones $X_1, X_2, \dots$, i.i.d., donde f, g son las densidades bajo H_0 y H_1 , respectivamente. También se trata el problema Kiefer-Weiss modificado, es decir, se minimiza $E(N)$ bajo una tercera densidad p que no necesariamente debe ser f o g . Además, no se tiene restricción alguna sobre el número de observaciones, bajo las probabilidades hipotetizadas por lo que $P(N = n) > 0$ para toda $n \in \mathbb{N}$, lo que significa que de manera implícita el horizonte del experimento es infinito; sin embargo, bajo la tercera densidad, p , $P(N < \infty) = 1$. La importancia de este trabajo radica básicamente en que se da una nueva demostración de la SPRT, utilizando el enfoque que se utiliza en el método de operadores de Lagrange (Rockafellar, 1970). Este método de Lagrange no es propiamente el mismo, pero se toma la misma idea del método de multiplicadores de Lagrange, porque a la función objetivo a minimizar, $E(N)$, se le suman las restricciones dadas por los errores tipo I y II. Es así que la función a minimizar ahora es $E(N) + u\alpha + v\beta$ (conocida como función de Lagrange), donde u, v son dos pesos dados a las cotas de las probabilidades de error tipo I y II, respectivamente. Un trabajo más reciente sobre el Lagrangiano para pruebas secuenciales con un enfoque de programación lineal lo pueden ver en Fauß y Zoubir (2015).

Hoy en día existen muchas aplicaciones de las pruebas de hipótesis secuenciales como podemos ver en [Bartroff et al. \(2013\)](#); [Dayanik y Yu \(2013\)](#); [Frazier y Yu \(2008\)](#); [Liu y Blöstein \(1992\)](#), entre otros. Precisamente y debido al análisis de ciertas investigaciones como las anteriores, nos surgen preguntas como, ¿qué sucedería si por alguna razón el experimento secuencial se ve interrumpido en algún momento por cuestiones ajenas a nuestra decisión? ¿Qué decisión tomar sobre las hipótesis? Por ejemplo, se puede pensar que se terminan los recursos para seguir tomando observaciones, o bien se precisa dar una decisión final antes de lo esperado, etc. Lo que queremos decir con lo antes expuesto, es que puede ocurrir que nuestro experimento se vea interrumpido por una variable independiente de las observaciones, y entonces, ¿qué decisión debemos tomar sobre nuestras hipótesis, ya sea tanto para el caso truncado como el no truncado? Lo que sigue es exponer brevemente algunos trabajos que son parte de la motivación principal de nuestra investigación.

Entonces continuando con los antecedentes, tenemos que en el trabajo de [Liu y Blöstein \(1992\)](#), se estudia la optimalidad de las SPRT's no estacionarias a través de un enfoque bayesiano, agregando un costo $c > 0$ fijo por observación y revisando el caso truncado. Sin embargo, en el caso truncado, si en la última observación tomada la razón de verosimilitud está en el intervalo de continuación, **no** se toma decisión alguna sobre las hipótesis. Para estudiar la optimalidad se toman observaciones independientes, pero no necesariamente idénticamente distribuidas, por lo que es más general este análisis. También se da la posibilidad de que el tiempo de paro, llamado N , sea aleatorio, pero dependiente de las observaciones, por ser dado por la regla de paro, que a su vez depende de las observaciones. Debido a este estudio, se obtienen ciertas propiedades que son aplicadas a distintos tipos de señales como: señales periódicas, transitorias y monótonas en ruido estacionario.

Es bien conocido que los resultados clásicos del análisis secuencial pueden ser obtenidos con base en la teoría de paro óptimo para *Procesos de decisión de Markov*, ver por ejemplo [Puterman \(1994\)](#) y [Cruz-Suárez et al. \(2014\)](#). En estos trabajos las pruebas secuenciales son llamadas políticas, en donde las observaciones se suponen i.i.d. y las cuales se analizan bajo un horizonte aleatorio independiente de las mismas observaciones. Este modelo de horizonte aleatorio e independiente de las observaciones se ajusta muy bien para modelar: fallas de un sistema en un modelo económico, la muerte de un animal en un modelo ecológico, la banca rota de una firma en un modelo económico, etc. Principalmente se emplea un horizonte aleatorio bajo una distribución geométrica.

El clásico “problema de la secretaria” para ocupar una vacante con las mejores aptitudes, ([Samuel-Cahn, 1996](#)), se modela bajo un horizonte aleatorio independiente de las observaciones i.i.d.. dicho horizonte se considera tanto finito como infinito. En este trabajo no se realiza prueba de hipótesis alguna, la decisión es sólo de tomar al candidato o a la candidata con mejores calificaciones para ocupar el puesto, por lo que el problema es puramente de paro óptimo.

Una de las motivaciones directa de nuestra investigación surge de los trabajos de [Frazier y Yu \(2008\)](#) y [Dayanik y Yu \(2013\)](#), en donde sí se aborda el problema de prueba de hipótesis secuencial simple cuando la decisión a tomar sobre una de las pruebas se debe

hacer bajo la presión de un horizonte aleatorio, independiente de las observaciones. En [Frazier y Yu \(2008\)](#) podemos ver que las pruebas de hipótesis son aplicadas al estudio de la neurociencia, en donde explican que aparentemente tanto los humanos como los animales acumulan información para posteriormente tomar, la que ellos denominan, decisión forzada óptima sobre dos alternativas simples (hipótesis nula y alternativa simples). Sin embargo, en muchas ocasiones la decisión final puede ser afectada por una fecha límite finita (horizonte aleatorio), que puede tener relación o no con la información acumulada, es decir, pueden ser factores externos o la misma incertidumbre del individuo. Realizan una simulación computarizada en donde presentan ejemplos con observaciones de Bernoulli, y suponen primero una fecha límite finita (caso truncado en el análisis clásico) y posteriormente emplean una distribución gamma para el horizonte aleatorio. Es importante aclarar que si la toma de decisión se ve afectada por el horizonte aleatorio, entonces se hace una penalización y la decisión final sobre una de las hipótesis no se lleva a cabo.

En cuanto al trabajo de [Dayanik y Yu \(2013\)](#), la fecha límite estocástica (horizonte aleatorio), como ellos denominan, y en diferencia del anterior trabajo, es que la decisión de la distribución de las observaciones se toma de un conjunto de m distintas funciones de densidad de probabilidad. Además, por el enfoque ahí empleado (aplicado a la neurociencia y psicología), se dan cuenta que minimizar el riesgo bayesiano es igual a maximizar la tasa de recompensa a largo plazo. Sin embargo, no se realiza decisión final sobre una de las hipótesis si el horizonte interrumpe el experimento y es entendible por el enfoque del área de aplicación. Es importante señalar que el horizonte aleatorio se supone independiente de las observaciones i.i.d y cuya su distribución suponen geométrica.

Un trabajo más reciente relacionado con esta investigación y bajo un enfoque bayesiano es propuesto en [Zhang et al. \(2016\)](#), en donde se supone un horizonte tanto aleatorio como finito para la toma de las observaciones independientes e idénticamente distribuidas. Se propone probar H_0 contra H_1 , minimizando el valor esperado muestral sólo bajo la hipótesis alterna. Se analiza el caso truncado, con N máximo de observaciones y la regla de paro está relacionada con la razón de verosimilitud y se acepta H_1 si la razón de verosimilitud supera una cota superior que se define de las mismas observaciones y que puede cambiar de etapa en etapa; sin embargo, si el tamaño muestral máximo se alcanza con la observación x_N , sin que la razón de verosimilitud supere la cota, entonces se decidirá sobre H_0 . En cuanto al caso en que el horizonte no es finito y en donde suponen las mismas condiciones que el caso truncado, proponen que sea un horizonte aleatorio, que denominan los autores tamaño muestral máximo, el cual es independiente de las observaciones y con distribución geométrica con media finita. Para tal propósito, se plantean pruebas secuenciales que tienen una regla de paro dada por las mismas observaciones, digamos T' y en donde actúa un horizonte aleatorio con distribución geométrica que es independiente de las observaciones, digamos N , y en tales pruebas secuenciales siempre se llegará a una decisión final. Por ejemplo si el experimento es interrumpido por el horizonte y no por la regla de paro de la prueba, entonces se decidirá por H_1 , mientras que si se detiene el experimento por la regla de paro de la prueba secuencial, entonces entra la regla de decisión definida en la prueba secuencial, por lo tanto, dependerá de la decisión optar por H_0 o bien H_1 .

Como antecedente final, tenemos las investigaciones de [Novikov \(2008\)](#) y [Novikov](#)

(2009), en donde caracteriza las pruebas secuenciales óptimas tanto para observaciones que sólo son independientes como para observaciones en general que provienen de un proceso estocástico a tiempo discreto, en donde minimiza el tamaño muestral esperado bajo una tercera hipótesis (Problema Kiefer-Weiss modificado) sobre todas las pruebas secuenciales cuyas probabilidades de errores tipo I y II permanecen acotadas y el enfoque del Lagrangiano también es aplicado.

Para nuestra presente investigación seguiremos en gran parte el enfoque dado por Novikov (2009), sólo que ahora participará un horizonte aleatorio cualquiera, independiente del proceso y con distribución conocida, en particular analizaremos ciertos resultados cuando el horizonte es geométrico, y a diferencia del caso truncado con Zhang et al. (2016), se tendrá una regla de paro dada por las observaciones y también el horizonte aleatorio que puede interrumpir la toma de observaciones antes de actuar la misma regla de paro o bien el tamaño muestral máximo permitido. Hasta donde sabemos no hay trabajos bajo este mismo enfoque, por lo tanto la teoría a desarrollar en esta investigación consideramos es original.

1.2. Conceptos básicos

En esta sección revisaremos algunos conceptos básicos que utilizaremos a lo largo de esta investigación.

Recordemos que en un espacio de probabilidad $(\mathcal{X}, \mathfrak{X}, P)$, $\mathfrak{X}$ es una σ -álgebra de eventos de un espacio muestral $\mathcal{X}$, la cual contiene toda la información de $\mathcal{X}$ y P es una medida de probabilidad (medida finita, por lo tanto, σ -finita) definida en $\mathfrak{X}$, i.e., $P : \mathfrak{X} \rightarrow [0, 1]$. Sea además, $\mathfrak{X}^n$ la colección de todos los eventos que suceden hasta el tiempo n . Entonces cada $\mathfrak{X}^n$ representa toda información disponible hasta el tiempo n . Por lo tanto:

- (i) para cada n se tiene que $\mathfrak{X}^n$ contiene menos información que $\mathfrak{X}$. Luego, $\mathfrak{X}^n \subset \mathfrak{X}$ para toda n ;
- (ii) y si $n < m$, entonces se tendrá mayor información conforme pasa el tiempo, por lo que $\mathfrak{X}^n \subset \mathfrak{X}^m$.

Decimos que $\{\mathfrak{X}^n : n = 0, 1, 2, \dots\}$ es una colección creciente de σ -álgebras. Así tenemos una estructura de la información, dependiente del tiempo, que filtra la información en $(\mathcal{X}, \mathfrak{X}, P)$. A toda estructura de la información que depende del tiempo (para nuestros fines n siempre es discreto) y que satisface los puntos (i) y (ii) se le conoce como filtración en $(\mathcal{X}, \mathfrak{X}, P)$. Con lo anterior se define un espacio filtrado como $(\mathcal{X}, \mathfrak{X}, \{\mathfrak{X}^n\}, P)$.

Se sabe además, que si la única información que se filtra proviene de una sucesión de variables aleatorias $X_1, X_2, X_3, \dots$, entonces decimos que la filtración $\{\mathfrak{X}^n\}$ es la filtración natural y se denota como

$$\sigma\{X_1, X_2, \dots, X_n\} =: \mathfrak{X}^n \quad \forall n.$$

Un proceso estocástico es una sucesión de variables aleatorias que cambian con el tiempo, donde el tiempo puede ser considerado continuo o discreto. En la teoría clásica

secuencial y debido a la naturaleza del problema, se consideran procesos estocásticos a tiempo discreto y es común el desarrollo de la misma bajo variables aleatorias continuas ya que para el caso discreto, y cuando es necesario, sólo se intercambia la integral por la suma.

Definimos a la siguiente sucesión $\{X_n\} = \{X_n(\omega) : n \in \mathbb{N}, \omega \in \mathcal{X}\}$ como un proceso estocástico a tiempo discreto adaptado en el espacio filtrado $(\mathcal{X}, \mathfrak{X}, \{\mathfrak{X}^n\}, P)$, donde $\{\mathfrak{X}^n\}$ es la filtración natural del proceso. Para el vector $(X_1, \dots, X_n)$ sea f_θ^n su función de densidad

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

dada por la derivada de Radon-Nikodym de su distribución, con respecto a una medida producto μ^n , donde μ es una medida σ -finita en $\mathfrak{X}$ para $n = 1, 2, 3, \dots$

Dentro de la teoría secuencial es imprescindible trabajar con tiempos de paro. Entonces un tiempo de paro es una función

$$\tau(\omega) : \mathcal{X}^\infty \rightarrow \{n : n \geq 0\}$$

donde $\mathcal{X}^\infty = \mathcal{X} \times \mathcal{X} \times \dots$, tal que el conjunto $\{\omega : \tau(\omega) \leq n\} =: \{\tau \leq n\}$ es medible con respecto a la filtración natural $\mathfrak{X}^n$ para toda n . En (Wald, 1947) vemos que los tiempos de paro son dados por la filtración natural de las mismas observaciones, pero también τ se trata de manera determinística al abordar el problema secuencial truncado. Recordemos además que si τ_1 y τ_2 son dos tiempos de paro definidos en el mismo espacio filtrado, o simplemente espacio, entonces, $\min\{\tau_1, \tau_2\}$ también es un tiempo de paro en dicho espacio.

Como realizaremos en un principio pruebas de hipótesis paramétricas, dadas ciertas observaciones $X_1, \dots, X_n$, entonces las pruebas que podemos hacer, pueden ser sobre un parámetro en particular, digamos $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$, o sobre la misma densidad de las n observaciones f_θ (f_θ conocida, no así θ), por ejemplo $H_0 : f_\theta = f_0$ vs $H_1 : f_\theta = f_1$, o bien sobre su distribución P_θ , así $H_0 : P_\theta = P_0$ vs $H_1 : P_\theta = P_1$. Lo que queremos decir con lo anterior, que en lo sucesivo puede aparecer alguna de estas pruebas de hipótesis y todas son equivalentes en el caso paramétrico.

Por el momento diremos que dada una sucesión de variables aleatorias $\{X_n\}$, una prueba secuencial es una pareja (ψ, ϕ) formada por una *regla de paro* ψ y una *regla de decisión* ϕ , dadas de la siguiente manera:

$$\psi = (\psi_0, \psi_1, \psi_2, \dots)$$

y

$$\phi = (\phi_0, \phi_1, \phi_2, \dots),$$

donde las funciones $\psi_n = \psi_n(X_1, \dots, X_n)$ y $\phi_n = \phi_n(X_1, \dots, X_n)$ son medibles con respecto a las σ -álgebras $\mathfrak{X}^n$, $n = 1, 2, \dots$, y $\psi_0, \phi_0 \in [0, 1]$ son dos números.

El análisis secuencial en el marco de un evento que puede interrumpir nuestro experimento (la toma de observaciones) por causas ajenas a nuestra propia decisión, lo llamaremos horizonte aleatorio y denotaremos por H . De esta manera, decimos que la variable aleatoria H es independiente de las observaciones $X_1, X_2, \dots, X_n$, pero supondremos en

todo momento que su distribución es conocida, y de hecho abordaremos más adelante el caso particular cuando su distribución es geométrica.

Debido a este horizonte aleatorio H el espacio filtrado sufre una ligera modificación. Por lo tanto si $(\mathcal{X}, \mathfrak{X}, \{\mathfrak{X}^n\}, P)$ es un espacio filtrado, entonces $\{\mathfrak{X}^n\}$ estará dada por

$$\mathfrak{X}^n = \sigma\{X_1, \dots, X_n, 1_{\{H=n\}}, 1_{\{H>n\}}\} \quad \text{para toda } n. \quad (1.3)$$

1.3. Planteamiento del problema

Supongamos que se desea realizar un experimento (estadístico) bajo un horizonte aleatorio H , que puede ser finito, $H < \infty$ o bien infinito $H \equiv \infty$, que es independiente del experimento en sí, pero con distribución conocida. Sea $\{X_n : n \geq 1\}$ un proceso estocástico a tiempo discreto adaptado en el espacio filtrado $(\mathcal{X}, \mathfrak{X}, \mathfrak{X}^n, P)$, donde $\mathfrak{X}^n$ es la filtración dada por proceso y el horizonte H como en (1.3), que es la información obtenida hasta el momento n debido a las observaciones $\{X_k : k = 1, \dots, n\}$ y por el conocimiento de si ha llegado o no a interrumpirse nuestro experimento debido a H al tiempo n .

Consideremos el problema clásico de probar una hipótesis simple a cerca de la distribución de las observaciones del proceso P , $H_0 : P = P_0$ contra una alternativa simple también $H_1 : P = P_1$.

Supongamos que, bajo la hipótesis H_i , el vector de observaciones $(X_1, X_2, \dots, X_n)$ tiene como “densidad” de probabilidad la función

$$f_i^n = f_i^n(x_1, x_2, \dots, x_n), \quad i = 0, 1,$$

dada por la derivada Radon-Nikodym de su distribución, con respecto a la medida producto μ^n , n veces μ por él mismo, donde μ es alguna medida σ -finita en $\mathfrak{X}$ para $n = 1, 2, 3, \dots$

Seguiremos con la idea de trabajar las pruebas secuenciales de forma aleatorizadas como [Ferguson \(1967\)](#), con lo que tenemos la siguiente.

Definición 1.1. *Una prueba de hipótesis secuencial (aleatorizada) es una pareja (ψ, ϕ) formada por una regla de paro ψ y una regla de decisión ϕ , con*

$$\psi = (\psi_0, \psi_1, \psi_2, \dots),$$

y

$$\phi = (\phi_0, \phi_1, \phi_2, \dots),$$

donde las funciones $\psi_n : \mathcal{X}^n \rightarrow [0, 1]$ y $\phi_n : \mathcal{X}^n \rightarrow [0, 1]$ suponemos medibles con respecto a las σ -álgebras $\mathfrak{X}^n$, $n = 1, 2, \dots$, y $\psi_0, \phi_0 \in [0, 1]$ son dos números.

Podemos debilitar la aleatoriedad a las reglas de paro y de decisión, siempre que la distribución de las observaciones del proceso sean continuas ([Fauß y Zoubir, 2015](#)).

Sabemos que en la teoría clásica [Wald y Wolfowitz \(1948\)](#) trabajaron con reglas de decisión indicadoras en la demostración de la optimalidad de la SPRT, debido a que dichas reglas podían tomar sólo valores cero o uno y facilitando con ello los cálculos. Además, en la teoría clásica es común tomar al menos una observación, i.e., $n = 1$, para iniciar el análisis secuencial.

En la definición anterior el valor ψ_0 es interpretado como la probabilidad de parar sin tomar observación alguna (que llamamos aleatorización inicial), y dado cualquier número n de observaciones, el valor $\psi_n(x_1, \dots, x_n)$ es interpretado como la probabilidad condicional de parar el experimento en la etapa n , dada las observaciones $(x_1, x_2, \dots, x_n)$, y dado que el experimento no paró antes. Estas dos reglas se aplican secuencialmente para $n = 0, 1, \dots$, hasta que el experimento finalmente se detenga.

Después de que el experimento se detuvo, la función de decisión ϕ entra en juego y es utilizada para tomar la decisión final: $\phi_n(x_1, x_2, \dots, x_n)$ se interpreta entonces como la probabilidad condicional de rechazar la hipótesis nula, dado que la información observada en la etapa n es $(x_1, x_2, \dots, x_n)$, y dado que el experimento termina en la etapa $n = 1, 2, \dots$. En caso de “detener” (de hecho nunca se inicia) el experimento sin tomar observación alguna, esto sucede en la etapa 0, entonces la hipótesis nula H_0 es rechazada con probabilidad ϕ_0 y será aceptada con probabilidad $1 - \phi_0$.

Por lo anterior es conveniente considerar a ψ_0 y ϕ_0 como funciones medibles en el conjunto $\mathcal{X}^0 = \{()\}$ (sin observaciones) con respecto a la σ -álgebra trivial $\mathfrak{X}^0 = \{\mathcal{X}^0, \emptyset\}$. Sea en este caso μ^0 una medida de probabilidad en $\mathfrak{X}^0$. Para mantener la coherencia, también definamos $f_0^0 = f_1^0 = 1$ (y, generalmente, $f^0 = 1$ en $\mathcal{X}^0$).

Notemos ahora que las pruebas de hipótesis clásicas donde la muestra tiene tamaño fijo n lo podemos ver como un caso particular de las pruebas de hipótesis secuenciales al tomar como regla de paro a $\psi = (\psi_0 \equiv 0, \psi_1 \equiv 0, \dots, \psi_{n-1} \equiv 0, \psi_n \equiv 1, \dots)$ y sólo se identifican con la decisión ϕ_n .

Cualquier regla de paro ψ nos genera, por el mismo proceso descrito anteriormente, un *tiempo de paro* τ_ψ (aleatorizado) con la siguiente distribución:

$$P(\tau_\psi = 0) = \psi_0,$$

y

$$P(\tau_\psi = n) = E[(1 - \psi_0)(1 - \psi_1) \cdots (1 - \psi_{n-1})\psi_n], \quad n = 1, 2, \dots \quad (1.4)$$

Ahora bien, para toda regla de paro $\psi = (\psi_0, \psi_1, \dots)$, denotemos a $t_0^\psi = 1$, $s_0^\psi = \psi_0$, y para cualquier $n = 1, 2, \dots$

$$t_n^\psi = (1 - \psi_0)(1 - \psi_1) \cdots (1 - \psi_{n-1}), \quad s_n^\psi = t_n^\psi \psi_n.$$

En particular, tenemos de (1.4):

$$P(\tau_\psi = n) = E s_n^\psi, \quad n = 0, 1, \dots$$

Observemos que para $n = 1, 2, 3, \dots$, la función $s_n^\psi = s_n^\psi(X_1, \dots, X_n)$ nos señalará el momento en que el experimento se detendrá exclusivamente por nuestra regla de paro en la etapa n , y no antes y ni debido al horizonte H . Mientras que la función $t_n^\psi = t_n^\psi(X_1, \dots, X_n)$ nos señalará el momento de detener el experimento en la etapa n debido al horizonte aleatorio, y no por la naturaleza del experimento. No debe haber confusión al utilizar esta P , que representa la probabilidad bajo la distribución verdadera de las observaciones, que bien puede ser P_0 , P_1 o simplemente P .

El enfoque principal de esta investigación es introducir un horizonte aleatorio H , donde la variable aleatoria H es una variable que toma valores enteros (posiblemente extendida) tal que $P(0 \leq H \leq \infty) = 1$ o de manera equivalente que $P(H = \infty) = 0$. Supondremos también que H es independiente del proceso de observaciones $X_1, X_2, \dots$, y que la distribución de la misma es fija y conocida por el estadista, a quien sólo se le informa si $H > n$ o no. En caso de que se interrumpa la toma de observaciones por el horizonte H se deberá realizar la decisión final en favor de H_0 o H_1 con base en la realización de las observaciones $(x_1, x_2, \dots, x_n)$, $n = 1, 2, \dots$ con que se cuente hasta ese momento. Ahora bien, cuando el horizonte es finito, $H \equiv N$ (caso truncado), para alguna constante N , el marco teórico correspondiente para abordar este problema son las pruebas secuenciales truncadas (Wald, 1947), con un número máximo de observaciones N , sin embargo, nosotros supondremos en el caso truncado que también puede actuar H de manera aleatoria.

Sabemos que al realizar una prueba de hipótesis no quedamos exentos de tomar una decisión errónea; por ejemplo, optar por rechazar la hipótesis nula H_0 (error del tipo I) cuando ésta es verdadera o aceptar la hipótesis nula, siendo falsa (error del tipo II). A las probabilidades de los errores tipo I y tipo II, se les denota comúnmente como α y β , respectivamente. Lo más sensato es tener estos errores α y β lo más pequeños posibles para tener la mejor decisión. Ahora bien, dada una prueba secuencial (ψ, ϕ) le asociaremos sus probabilidades de error tipo I y tipo II como $\alpha(\psi, \phi)$ y $\beta(\psi, \phi)$, respectivamente.

Notemos que dada una prueba secuencial y bajo el marco de un horizonte aleatorio H , $\alpha(\psi, \phi)$ y $\beta(\psi, \phi)$ se pueden calcular partiendo el espacio donde está definido el horizonte aleatorio, $\mathbb{Z}^+$ en dos partes, tomando una parte cuando H llega antes o igual al tiempo de paro y la segunda parte cuando el tiempo de paro llega antes que el horizonte aleatorio. Por lo tanto, las probabilidades de error tipo I y tipo II pueden calcularse de la siguiente manera:

$$\begin{aligned}
 \alpha(\psi, \phi) &= P_0(\text{rechazar } H_0, H > \tau_\psi) + P_0(\text{rechazar } H_0, H \leq \tau_\psi) \\
 &= \sum_{n=0}^{\infty} P_0(\phi_n = 1, s_n^\psi = 1, H > n) + \sum_{n=0}^{\infty} P_0(\phi_n = 1, t_n^\psi = 1, H = n) \\
 &= \sum_{n=0}^{\infty} E_0 \phi_n s_n^\psi P(H > n) + \sum_{n=0}^{\infty} E_0 \phi_n t_n^\psi P(H = n) \\
 &= \sum_{n=0}^{\infty} E_0 \phi_n t_n^\psi (\psi_n P(H > n) + P(H = n)) \quad (\text{agrupando y factorizando } t_n^\psi)
 \end{aligned}$$

y

$$\begin{aligned}
 \beta(\psi, \phi) &= P_1(\text{aceptar } H_0, H > \tau_\psi) + P_1(\text{aceptar } H_0, H \leq \tau_\psi) \\
 &= \sum_{n=0}^{\infty} P_1(1 - \phi_n = 1, s_n^\psi = 1, H > n) + \sum_{n=0}^{\infty} P_1(1 - \phi_n = 1, t_n^\psi = 1, H = n) \\
 &= \sum_{n=0}^{\infty} E_1(1 - \phi_n) s_n^\psi P(H > n) + \sum_{n=0}^{\infty} E_1(1 - \phi_n) t_n^\psi P(H = n) \\
 &= \sum_{n=0}^{\infty} E_1(1 - \phi_n) t_n^\psi (\psi_n P(H > n) + P(H = n)).
 \end{aligned}$$

De esta manera tenemos la siguiente.

Definición 1.2. Dada cualquier prueba secuencial (ψ, ϕ) , definimos las probabilidades de error tipo I como

$$\alpha(\psi, \phi) = \sum_{n=0}^{\infty} E_0 \phi_n t_n^\psi h_n^\psi \quad (1.5)$$

y tipo II como

$$\beta(\psi, \phi) = \sum_{n=0}^{\infty} E_1 (1 - \phi_n) t_n^\psi h_n^\psi, \quad (1.6)$$

donde $h_n^\psi = \psi_n P(H > n) + P(H = n)$ para toda n .

Enseguida veremos qué sucede con el *número promedio muestral*, el cual debe tener en cuenta las restricciones debido a las mismas observaciones y al horizonte aleatorio H ; por lo tanto, estamos interesados en el valor esperado de $\tau_\psi \wedge H = \min\{\tau_\psi, H\}$ bajo alguna de las dos hipótesis en cuestión, H_i , $i = 0, 1$ y el cual denotaremos como:

$$\mathcal{N}(\psi) = E(\tau_\psi \wedge H). \quad (1.7)$$

Supongamos que la distribución verdadera de las observaciones es $P = P_i$, $i = 0, 1$. Entonces si la probabilidad de terminar el experimento, ya sea por el tiempo de paro definido en la prueba secuencial o debido al horizonte aleatorio, es finita e igual a uno, $P(\tau_\psi \wedge H < \infty) = 1$ (equivalentemente $P(\tau_\psi \wedge H = \infty) = 0$), se sigue que

$$\begin{aligned} \mathcal{N}(\psi) &= E\tau_\psi \wedge H I_{\{H \geq \tau_\psi\}} + E\tau_\psi \wedge H I_{\{H < \tau_\psi\}} \\ &= E\tau_\psi I_{\{H \geq \tau_\psi\}} + E H I_{\{H < \tau_\psi\}} \\ &= \sum_{n=0}^{\infty} n P(\tau_\psi = n, H > n) + \sum_{n=0}^{\infty} n P(H = n, \tau_\psi > n) \\ &= \sum_{n=0}^{\infty} n P(\tau_\psi = n) P(H > n) + \sum_{n=0}^{\infty} n P(H = n) P(\tau_\psi > n) \\ &= \sum_{n=0}^{\infty} n E s_n^\psi P(H > n) + \sum_{n=0}^{\infty} n P(H = n) E t_n^\psi \\ &= \sum_{n=0}^{\infty} n E (s_n^\psi P(H > n) + t_n^\psi P(H = n)) = \sum_{n=0}^{\infty} n E t_n^\psi h_n^\psi \end{aligned}$$

donde I_A es la función indicadora del conjunto A .

Definición 1.3. Bajo un horizonte aleatorio, si $P(\tau_\psi \wedge H < \infty) = 1$, tenemos que el valor esperado del mínimo entre el tiempo de paro τ_ψ y el horizonte aleatorio H es

$$\mathcal{N}(\psi) = \sum_{n=0}^{\infty} n E t_n^\psi h_n^\psi. \quad (1.8)$$

De otra manera si $P(\tau_\psi \wedge H < \infty) < 1$ (o equivalentemente si $P(\tau_\psi \wedge H = \infty) > 0$), entonces $\mathcal{N}(\psi) = \infty$.

En un contexto usual de pruebas de hipótesis secuenciales, es conveniente trabajar con pruebas para las cuales sus errores tipo I y tipo II no excedan ciertas cotas que imponemos de antemano, digamos:

$$\alpha(\psi, \phi) \leq \alpha \quad \text{y} \quad \beta(\psi, \phi) \leq \beta, \quad (1.9)$$

para algunas $\alpha, \beta \in [0, 1)$ fijas.

Entonces, bajo los conceptos antes planteados, el problema principal es caracterizar las pruebas secuenciales óptimas que surgen de minimizar el tamaño muestral esperado $\mathcal{N}(\psi)$ sobre todas las pruebas secuenciales sujetas a (1.9). Supondremos que la esperanza, definida en (1.7), es calculada bajo una tercera distribución de las observaciones del proceso $X_1, X_2, \dots$, que no necesariamente tiene que coincidir con P_0 o P_1 , i.e. consideraremos una variante del problema Kiefer-Weiss modificado (ver [Kiefer y Weiss \(1957\)](#), [Weiss \(1962\)](#), [Lorden \(1980\)](#) con observaciones i.i.d.), pero bajo horizonte aleatorio (KWMH).

Capítulo 2

Pruebas Óptimas bajo Horizonte

2.1. Reducción a un problema de paro óptimo

Primero abordaremos el problema Kiefer-Weiss modificado bajo horizonte aleatorio (KWMH) en la clase Δ de todas las pruebas secuenciales. En realidad Δ es una clase genérica de las pruebas secuenciales en la cual requerimos obtener la minimización, por lo que Δ puede ser tomada como la clase de todas las pruebas secuenciales truncadas o bien puede ser tomada simplemente como la clase de todas las pruebas secuenciales. La idea de este capítulo radica primero en abordar el problema truncado, para luego extender los resultados para el caso no truncado.

Para minimizar (1.8) sujeto a (1.9) aplicaremos el método de multiplicadores de Lagrange usado por Lorden (1980) con el objeto de reducir el problema de minimización de la función de Lagrange:

$$\mathcal{N}(\psi) + u\alpha(\psi, \phi) + v\beta(\psi, \phi), \quad (2.1)$$

donde $u, v > 0$ son ciertas constantes llamadas multiplicadores de Lagrange.

Es inmediato ver que minimizar (2.1) es equivalente a minimizar

$$\mathcal{L}(\psi, \phi) = \mathcal{L}(\psi, \phi; c, \lambda) = c\mathcal{N}(\psi) + \lambda\alpha(\psi, \phi) + \beta(\psi, \phi), \quad (2.2)$$

donde $c, \lambda > 0$ son algunas constantes (c es el costo unitario por observación); de hecho, $c = 1/v$ y $\lambda = u/v$. Usaremos esta última forma, la cual es más conveniente para nuestros propósitos.

Las constantes de peso u, v , también llamadas multiplicadores de Lagrange, han recibido poca atención dentro de la literatura en cuanto a cómo tomarlas de tal manera que satisfaga las restricciones de las probabilidades en (1.9), sin embargo, pueden consultar Fauß y Zoubir (2015) donde se adentran al estudio de las mismas. Precisamente en el trabajo de Fauß y Zoubir (2015) hacen uso del método de Lagrange para pruebas de hipótesis secuenciales bajo un enfoque de programación lineal.

Observemos que a partir de las definiciones 1.2 y 1.3 tenemos que la función $\mathcal{L}(\psi, \phi)$ (nombrado Lagrangiano en alguna literatura) dada en (2.2) a minimizar es:

$$\mathcal{L}(\psi, \phi) = \sum_{n=0}^{\infty} \int t_n^\psi h_n^\psi (cnf^n + \lambda\phi_n f_0^n + (1 - \phi_n)f_1^n) d\mu^n. \quad (2.3)$$

El siguiente lema nos servirá para fijar el camino riguroso a nuestro problema de minimizar $\mathcal{N}(\psi)$ que, bajo las restricciones (1.9), se reduce a la minimización de $\mathcal{L}(\psi, \phi)$, de hecho es el Teorema 2.1 en Novikov (2009) (simplemente hay que recordar que el horizonte aleatorio H se encuentra detrás de todas las definiciones de ahora en adelante).

Lema 2.1. *Sean c y λ algunas constantes positivas. Sea $(\psi, \phi) \in \Delta$ una prueba que satisface $\mathcal{L}(\psi, \phi; c, \lambda) < \infty$, y tal que para toda prueba $(\psi', \phi') \in \Delta$ que cumpla*

$$\mathcal{L}(\psi, \phi; c, \lambda) \leq \mathcal{L}(\psi', \phi'; c, \lambda), \quad (2.4)$$

y que

$$\alpha(\psi, \phi) = \alpha \quad y \quad \beta(\psi, \phi) = \beta. \quad (2.5)$$

Entonces para toda prueba $(\psi', \phi') \in \Delta$ que satisface

$$\alpha(\psi', \phi') \leq \alpha \quad y \quad \beta(\psi', \phi') \leq \beta \quad (2.6)$$

se cumple que

$$\mathcal{N}(\psi) \leq \mathcal{N}(\psi'). \quad (2.7)$$

La desigualdad en (2.7) es estricta si al menos una de las desigualdades en (2.6) es estricta.

La prueba del Lema 2.1 se repite literalmente como en Novikov (2009).

Cabe mencionar que las condiciones (2.5) son en general difícil de cumplir debido a las constantes c y λ introducidas en (2.2), ya que no hay una manera simple de moverlas de tal forma que se cumpla (2.6). La pregunta sobre la existencia de “buenas” pruebas con α y β dadas que cumpla con (2.5), no hay una respuesta en general, inclusive para la SPRT (Wald y Wolfowitz, 1948). Por otro lado, en algún sentido, las condiciones (2.5) se satisfacen “automáticamente” cada vez que se tenga una prueba (ψ, ϕ) que cumpla (2.4); de hecho, podemos elegir $\alpha = \alpha(\psi, \phi)$ y $\beta = \beta(\psi, \phi)$ y (2.5) se cumple.

Debido al lema arriba descrito, el problema ahora se reduce a encontrar la estructura de las pruebas que minimizan la función $\mathcal{L}(\psi, \phi)$ sobre todas las pruebas secuenciales (ver condición (2.4)). En una manera muy usual, el primer paso es determinar las reglas de decisión óptimas que minimizan la función de Lagrange $\mathcal{L}(\psi, \phi)$ sobre todas las reglas de decisión ϕ . Precisamente el Teorema 2.1 abajo enunciado es el que nos dará las condiciones necesarias y suficientes para este propósito.

Antes necesitamos adoptar la siguiente notación para facilitar los cálculos. Escribamos $\psi \simeq I_{\{F_1 \leq F_2\}}$ cuando se cumpla la desigualdad siguiente $I_{\{F_1 < F_2\}} \leq \psi \leq I_{\{F_1 \leq F_2\}}$. Veamos que la notación antes descrita es equivalente a tener:

$$\psi = 1 \text{ si } F_1 < F_2,$$

$$\psi = 0 \text{ si } F_1 > F_2 \text{ y}$$

$0 \leq \psi \leq 1$ siempre que $F_1 = F_2$.

Si las funciones ψ , F_1 y F_2 dependen de algunos argumentos, entonces esta regla deberá aplicarse a sus valores calculados en cualquier argumento dado.

Acordemos a partir de ahora que para cualquier par de funciones $F_n = F_n(x_1, \dots, x_n)$ y $G_n = G_n(x_1, \dots, x_n)$ las relaciones $F_n = G_n$ y $F_n \leq G_n$, así también los límites como $\lim_{N \rightarrow \infty} F_n^N$, etc., se entenderá μ^n -casi en todas partes, abreviadamente como μ^n -c.t.p., a menos que se diga expresamente lo contrario. Vale la pena recordar que una condición se cumple μ^n -c.t.p. si existe un conjunto $A \in \mathfrak{X}^n$ donde la condición no se cumple y tiene medida cero, $\mu^n(A) = 0$, para alguna n .

Ahora es preciso ver el siguiente lema que es de suma importancia para la demostración del teorema posterior.

Lema 2.2. Sean ϕ, F_1, F_2 algunas funciones medibles en un espacio medible con una medida μ , tales que

$$0 \leq \phi(x) \leq 1, \quad F_1(x) \geq 0, \quad F_2(x) \geq 0$$

y

$$\int \min\{F_1(x), F_2(x)\} d\mu(x) < \infty.$$

Entonces

$$\int (\phi(x)F_1(x) + (1 - \phi(x))F_2(x)) d\mu(x) \geq \int \min\{F_1(x), F_2(x)\} d\mu(x), \quad (2.8)$$

con igualdad si y sólo si

$$\phi \simeq I_{\{F_1(x) \leq F_2(x)\}} \quad (2.9)$$

μ -c.t.p.

La prueba de este lema se encuentra en [Novikov \(2009, p.211\)](#).

Teorema 2.1. Sea (ψ, ϕ) cualquier prueba secuencial y λ un número positivo dado. Entonces

$$\lambda\alpha(\psi, \phi) + \beta(\psi, \phi) \geq \sum_{n=0}^{\infty} \int t_n^\psi h_n^\psi \min\{\lambda f_0^n, f_1^n\} d\mu^n, \quad (2.10)$$

donde

$$h_n^\psi = \psi_n P(H > n) + P(H = n), \quad n = 0, 1, 2, \dots \quad (2.11)$$

Hay igualdad en (2.10) si y sólo si la regla de decisión $\phi = (\phi_0, \phi_1, \dots)$ es tal que

$$\phi_n \simeq I_{\{\lambda f_0^n \leq f_1^n\}} \quad \mu^n\text{-c.t.p. en } S_n^\psi \quad (2.12)$$

para toda $n = 0, 1, \dots$, donde

$$S_n^\psi = \{(x_1, \dots, x_n) \in \mathcal{X}^n : t_n^\psi(x_1, \dots, x_n) h_n^\psi(x_1, \dots, x_n) > 0\}.$$

La demostración de este teorema la dejamos hasta el apéndice A.1.

Notemos que debido a nuestro previo acuerdo, para cuando no tomamos observación alguna, tenemos

$$S_0^\psi = \begin{cases} \mathcal{X}^0 & \text{si } \psi_0 P(H > 0) + P(H = 0) > 0, \text{ y} \\ \emptyset, & \text{en otro caso.} \end{cases} \quad (2.13)$$

En el primer caso, la condición “ $\phi_0 \simeq I_{\{\lambda f_0^0 \leq f_1^0\}}$ μ^0 -c.t.p. en S_0^ψ ” significa justamente que “ $\phi_0 \simeq I_{\{\lambda \leq 1\}}$ ”, en el segundo caso, no representa ninguna condición en absoluto.

Observemos que del Teorema 2.1 el $\inf_\phi \mathcal{L}(\psi, \phi)$ siempre se alcanza para una ϕ , cualquiera que sea la regla de paro ψ ; de hecho, podemos utilizar una única ϕ para toda ψ que cumple, por ejemplo, si las componentes son de la forma $\phi_n = I_{\{\lambda f_0^n \leq f_1^n\}}$, $n = 0, 1, \dots$. Con base en lo anterior y abusando de la notación, definamos

$$\mathcal{L}(\psi) = \mathcal{L}(\psi; c, \lambda) = \inf_\phi \mathcal{L}(\psi, \phi; c, \lambda). \quad (2.14)$$

Con lo anterior vemos que nuestro problema de minimización de $\mathcal{L}(\psi, \phi)$ se reduce a la minimización ahora de $\mathcal{L}(\psi)$ sobre todas las reglas de paro ψ . De hecho si existe tal regla de paro ψ , tal que

$$\mathcal{L}(\psi) = \inf_{\psi'} \mathcal{L}(\psi'; c, \lambda),$$

entonces usando cualquier regla ϕ que satisfaga (2.12) del Teorema 2.1 tendremos

$$\mathcal{L}(\psi, \phi) = \mathcal{L}(\psi) \leq \mathcal{L}(\psi') = \mathcal{L}(\psi', \phi) \leq \mathcal{L}(\psi', \phi'),$$

para cualquier prueba secuencial (ψ', ϕ') , que garantice la desigualdad (2.4) en el Lema 2.1.

Por lo tanto, y de ahora en adelante, el objetivo se centra en minimizar $\mathcal{L}(\psi)$ sobre todas las reglas de paro ψ .

Definamos a

$$l_n = l_n(x_1, \dots, x_n; \lambda) = \min\{\lambda f_0^n(x_1, \dots, x_n), f_1^n(x_1, \dots, x_n)\}$$

para $n = 0, 1, \dots$ (notemos que por lo ya convenido $f_0^0 = f_1^0 = 1$, luego $l_0 = \min\{\lambda, 1\}$). En el caso del problema KWMH, estamos suponiendo que la distribución P tiene función de densidad f^n dado un vector de observaciones del proceso, para toda n .

Corolario 2.1. Si $P(\tau_\psi \wedge H < \infty) = 1$, entonces

$$\mathcal{L}(\psi) = \sum_{n=0}^{\infty} \int t_n^\psi h_n^\psi (c_n f^n + l_n) d\mu^n. \quad (2.15)$$

Por otro lado si $P(\tau_\psi \wedge H < \infty) < 1$, entonces $L(\psi) = \infty$.

Demostración. Se sigue del Teorema 2.1 por (2.2), en vista de (1.8). □

2.2. Reglas de paro óptimas

En esta sección se centra el principal estudio de esta investigación ya que vamos a caracterizar las reglas óptimas, en caso truncado y no truncado. Primero abordaremos el problema de paro óptimo visto en la sección anterior.

2.2.1. Reglas de Paro Óptimo Truncadas

Decimos que una regla de paro es truncada al nivel N si no se permite tomar más de $N \geq 0$ observaciones del proceso. Podemos ver que una regla de paro $\psi = (\psi_0, \psi_1, \psi_2, \dots)$ será truncada al nivel $N \in \mathbb{N}$ si

$$(1 - \psi_0)(1 - \psi_1) \dots (1 - \psi_N) \equiv 0, \quad (2.16)$$

o de igual forma cuando $P(\tau_\psi = N) = 1$.

Sea $\mathcal{D}^N$ la clase de las reglas truncadas.

En lo siguiente caracterizaremos la estructura de las reglas en $\mathcal{D}^N$ tales que minimizan $\mathcal{L}(\psi)$, para cualquier $N \geq 0$ fija.

Observemos primero que, debido a (2.16) y (1.8), todos los sumandos con $n = N + 1, N + 2, \dots$ en la parte derecha de (2.15) son iguales a 0, luego

$$\mathcal{L}(\psi) = \sum_{n=0}^N \int t_n^\psi h_n^\psi (cnf^n + l_n) d\mu^n. \quad (2.17)$$

Además, el último sumando de la suma en (2.17) es

$$\begin{aligned} \int t_N^\psi h_N^\psi (cNf^N + l_N) d\mu^N &= P(H > N) \int t_N^\psi \psi_N (cNf^N + l_N) d\mu^N \\ &+ P(H = N) \int t_N^\psi (cNf^N + l_N) d\mu^N = P(H \geq N) \int t_N^\psi (cNf^N + l_N) d\mu^N, \end{aligned}$$

lo anterior es cierto debido a (2.16).

De esta manera tenemos que (2.17) se convierte en

$$\mathcal{L}_N(\psi) = \sum_{n=0}^{N-1} \int t_n^\psi h_n^\psi (cnf^n + l_n) d\mu^n + P(H \geq N) \int t_N^\psi (cNf^N + l_N) d\mu^N. \quad (2.18)$$

Tomemos en cuenta que si $P(H \geq N) = 0$, entonces la expresión en (2.18) se reduce a lo mismo para una N menor. De hecho podemos tomar a $N = \max\{n \geq 0 : P(H \geq n) > 0\}$. Por lo tanto, y sin perder generalidad, podemos suponer que se cumple $P(H \geq N) > 0$.

En vista de (2.18) definamos para cualquier función medible no negativa $v = v(x_1, \dots, x_k)$, $k \geq 1$, el funcional

$$B_k^\psi(v) = \sum_{n=0}^{k-1} \int t_n^\psi h_n^\psi (cnf^n + l_n) d\mu^n + P(H \geq k) \int t_k^\psi (ckf^k + v) d\mu^k \quad (2.19)$$

(de ahora en adelante, y para tener sentido, tomaremos a $\sum_{n=0}^{-1}(\cdot) = 0$).

Debido a las definiciones (2.18) y (2.19) se cumple la siguiente igualdad

$$\mathcal{L}_N(\psi) = B_N^\psi(l_N)$$

para toda $\psi \in \mathcal{D}^N$.

El siguiente lema representa la mayor parte del trabajo en esta sección.

Lema 2.3. *Sea $v = v(x_1, \dots, x_{k+1})$ cualquier función no negativa μ^{k+1} -integrable, donde $k \geq 0$ es tal que $P(H \geq k+1) > 0$. Entonces*

$$B_{k+1}^\psi(v) \geq B_k^\psi(D_k v), \quad (2.20)$$

donde, por definición,

$$D_k v = p_k l_k + (1 - p_k) \min \{l_k, c f^k + J_k v\},$$

con

$$J_k v = J_k v(x_1, \dots, x_k) = \int v(x_1, \dots, x_k, x_{k+1}) d\mu(x_{k+1})$$

y

$$p_k = \frac{P(H = k)}{P(H \geq k)}. \quad (2.21)$$

Se cumplirá la igualdad en (2.20) si y sólo si

$$\psi_k \simeq I_{\{l_k \leq c f^k + J_k v\}} \quad \mu^k\text{-c.t.p. en } T_k^\psi, \quad (2.22)$$

donde $T_k^\psi = \{(x_1, \dots, x_k) \in \mathcal{X}^k : t_k^\psi(x_1, \dots, x_k) > 0\}$.

La demostración del Lema 2.3 se encuentra en el Apéndice A.2.

Sea ahora

$$V_N^N = V_N^N(x_1, \dots, x_N) = l_N(x_1, \dots, x_N) \quad (2.23)$$

y, recursivamente para $n = N, N-1, \dots, 1$, definamos

$$V_{n-1}^N = D_{n-1} V_n^N. \quad (2.24)$$

Dada cualquier regla de paro truncada $\psi \in \mathcal{D}^N$, podemos aplicar el Lema 2.3 con $k = N-1$ para $\mathcal{L}_N(\psi) = B_N^\psi(l_N)$ con lo que obtenemos $\mathcal{L}_N(\psi) \geq B_{N-1}^\psi(V_{N-1}^N)$, con igualdad si y sólo si ψ_{N-1} cumple con (2.22). Ahora, aplicando Lema 2.3 con $k = N-2$ para $B_{N-1}^\psi(V_{N-1}^N)$ tenemos $B_{N-1}^\psi(V_{N-1}^N) \geq B_{N-2}^\psi(V_{N-2}^N)$, y con igualdad si y sólo si ψ_{N-2} cumple con (2.22). Haciendo esto de manera consecutiva para los demás $k = N-3, \dots, 1$, llegamos finalmente a

$$B_1^\psi(V_1^N) \geq B_0^\psi(V_0^N) = V_0^N = p_0 l_0 + (1 - p_0) \min \{l_0, c + J_0 V_1^N\}$$

y obviamente una serie de condiciones suficientes y necesarias sobre cada ψ_k para alcanzar las igualdades en todas las desigualdades. Entonces por una variante de inducción matemática hacia atrás (ver Chow et al. (1971)) obtenemos el siguiente.

Teorema 2.2. *Supongamos que $N \geq 1$ y que $P(H \geq N) > 0$. Entonces para toda $\psi \in \mathcal{D}^N$*

$$\mathcal{L}_N(\psi) \geq V_0^N. \quad (2.25)$$

Se cumplirá la igualdad en (2.25), para una $\psi \in \mathcal{D}^N$, si y sólo si

$$\psi_n \simeq I_{\{l_n \leq cf^n + J_n V_{n+1}^N\}} \quad \mu^n\text{-c.t.p. en } T_n^\psi$$

para toda $n = 0, 1, \dots, N-1$.

Conforme a nuestro previo acuerdo, tenemos que $T_0^\psi = \mathcal{X}^0$ es el conjunto de “no observaciones” y “ μ^0 -c.t.p. en el conjunto T_0^ψ ” significa justamente “en $\mathcal{X}^0$ ”.

Ahora es más fácil ver el caso cuando $P(H \geq N) = 0$. Para ello podemos aplicar el mismo procedimiento del Teorema 2.2, con una N inferior, como mencionamos anteriormente, con $N = \max\{n \geq 0 : P(H \geq n) > 0\}$. Esto precisamente nos lleva a que el horizonte aleatorio H es acotado con probabilidad uno, digamos $P(H \leq k) = 1$, y $N > k$; donde los componentes de la regla de paro $\psi_{k+1}, \psi_{k+2}, \dots, \psi_N$ de cualquier regla de paro ψ serán irrelevantes para llegar a la optimalidad en $\mathcal{D}^N$ debido a que no se requerirán nunca para el procedimiento de paro.

En caso de tener un truncamiento al nivel $N = 0$, existe una única regla de paro $\psi_0 = 1$ en $\mathcal{D}^0$, lo que significa tener la desigualdad $l_0 < cf^0 + J_0 V_1^0$ debido a (2.22). Ahora por (2.13), vemos que $S_0^\psi = \mathcal{X}^0$, entonces como regla de decisión tenemos que $\phi_0 \simeq I_{\{\lambda, 1\}} \mu^0$ -c.t.p. en S_0^ψ . Por lo tanto, tenemos que $\mathcal{L}_0(\psi_0) = V_0^0 = \min\{\lambda, 1\}$, entonces la minimización de $\mathcal{L}_0(\psi)$ sobre $\psi \in \mathcal{D}^0$ se vuelve trivial.

También se vuelve trivial la optimización si en alguna etapa $n < H$, una de las densidades conjuntas de las hipótesis se anula. Por ejemplo, si $f_1^n = 0$, para alguna n , entonces $l_n = \min\{\lambda f_0^n, 0\} = 0$. Luego, para alcanzar la cota inferior (2.25) del Teorema 2.2, bajo los mismos supuestos, $\psi_n = 1$, mientras que para $k = n-1, n-2, \dots, 1$, ψ_k es de la forma (2.22) y el elemento de la regla de decisión para la etapa n es $\phi_n \simeq I_{\{\lambda f_0^n \leq 0\}} = 0$; es decir, debemos rechazar la hipótesis nula con una probabilidad de 0. Por lo tanto, aceptamos la hipótesis nula es este caso y el costo asociado al experimento es cn .

Recordemos que las funciones V_n^N son construidas para cualquier c y λ positivas y fijas, así las funciones $V_n^N = V_n^N(x_1, \dots, x_n; c, \lambda)$ dependerán en cierta medida de dichas constantes. Además, la distribución del horizonte aleatorio H está detrás de todas nuestras definiciones y construcciones, por lo que es necesario tenerlo presente en lo sucesivo.

Cabe hacer mención que la variable p_k definida en (2.21) se le conoce como “tasa de fallos” y se puede demostrar, de manera sencilla, que cuando esta tasa es constante, en el caso discreto, tiene una distribución geométrica, mientras que para el caso continuo, tiene una distribución exponencial. Esta parte de la tasa de fallos la veremos con un poco de detalle más adelante cuando abordemos el caso de observaciones independientes e idénticamente distribuidas.

2.2.2. Optimalidad de las reglas de paro no truncadas

Ahora estamos en condiciones de extender los resultados previos para poder caracterizar la estructura de reglas de paro óptimas, cuando no existe restricción sobre el número

máximo de observaciones y aplicar estos resultados al problema original de pruebas de hipótesis. Sobre la regla de paro no hay restricción, sin embargo, recordemos que el horizonte aleatorio puede interrumpir el experimento en cualquier momento.

Dada cualquier regla de paro ψ y para cualquier número natural $N \geq 1$ definamos,

$$\mathcal{L}_N(\psi) = \mathcal{L}(\psi^N; c, \lambda),$$

donde $\psi^N = (\psi_0, \psi_1, \dots, \psi_{N-1}, 1, \dots)$ es una regla ψ truncada al nivel N . Como $\psi^N \in \mathcal{D}^N$, se sigue que $\mathcal{L}_N(\psi)$ puede ser calculado usando (2.18) y que la desigualdad en (2.25) se cumple para toda $N \geq 1$. Así obtendremos los resultados de nuestra optimalidad, precisamente pasando al límite en (2.25), como $N \rightarrow \infty$.

Como un primer paso para lo anterior, necesitamos saber las condiciones bajo las cuales se cumple $\mathcal{L}_N(\psi) \rightarrow \mathcal{L}(\psi)$, siempre que $N \rightarrow \infty$.

Lema 2.4. *Sean λ y c cualesquiera números positivos y ψ cualquier regla de paro tal que $E(\tau_\psi \wedge H) < \infty$. Entonces*

$$\lim_{N \rightarrow \infty} \mathcal{L}_N(\psi; c, \lambda) = \mathcal{L}(\psi; c, \lambda)$$

si y sólo si

$$\lim_{n \rightarrow \infty} P(H \geq n) \int t_n^\psi \min\{\lambda f_0^n, f_1^n\} d\mu^n = 0.$$

Si $E(\tau_\psi \wedge H) = \infty$ pero $P(\tau_\psi \wedge H < \infty) = 1$ entonces $\lim_{N \rightarrow \infty} \mathcal{L}_N(\psi) = \infty$.

La demostración del Lema 2.4 se encuentra en el Apéndice A.3.

Corolario 2.2. *Si $P(H < \infty) = 1$ entonces*

$$\lim_{N \rightarrow \infty} \mathcal{L}_N(\psi) = \mathcal{L}(\psi) \tag{2.26}$$

para toda regla de paro ψ .

Demostración. Si

$$E(\tau_\psi \wedge H) < \infty, \tag{2.27}$$

entonces (2.26) se sigue de la primera parte del Lema 2.4, porque

$$\int t_n^\psi \min\{\lambda f_0^n, f_1^n\} d\mu^n \leq \int \min\{\lambda f_0^n, f_1^n\} d\mu^n \leq \lambda \int f_0^n d\mu^n = \lambda < \infty,$$

y $P(H \geq n) \rightarrow 0$, $n \rightarrow \infty$, por condición.

Si (2.27) no se cumple, entonces $\mathcal{L}(\psi) = \infty$, y (2.26) se sigue de la segunda parte del Lema 2.4 debido a que $P(\tau_\psi \wedge H < \infty) \geq P(H < \infty) = 1$. $\square$

Corolario 2.3. *Supongamos que al menos una de las siguientes condiciones se satisface:*

- (i) $P = \pi_0 P_0 + \pi_1 P_1$ con $\pi_0 \geq 0$, $\pi_1 \geq 0$, donde $\pi_0 + \pi_1 = 1$.

(ii) para alguna $\lambda > 0$

$$\lim_{n \rightarrow \infty} \int \min\{\lambda f_0^n, f_1^n\} d\mu^n = 0.$$

(iii)

$$\frac{f_1^n(X_1, \dots, X_n)}{f_0^n(X_1, \dots, X_n)} \rightarrow 0, \quad \text{como } n \rightarrow \infty,$$

en P_0 -probabilidad.

Entonces para toda regla de paro ψ tal que $P(\tau_\psi \wedge H < \infty) = 1$ se cumple

$$\lim_{N \rightarrow \infty} \mathcal{L}_N(\psi) = \mathcal{L}(\psi). \quad (2.28)$$

La prueba del Corolario 2.3 la dejamos hasta el Apéndice A.4.

De igual manera necesitamos ver qué ocurre cuando $N \rightarrow \infty$ en las funciones (2.23) y (2.24).

Lema 2.5. Para todos los enteros no negativos n y N tales que $n \leq N$, se cumple

$$V_n^N \geq V_n^{N+1} \quad \mu^n\text{-c.t.p.} \quad (2.29)$$

Ver prueba en el apéndice A.5

Dado un número $N \geq 0$, sabemos que las funciones V_n^N , $0 \leq n \leq N$, son no negativas y por el Lema 2.5 sabemos que son no crecientes; por lo tanto, su límite existe cuando $N \rightarrow \infty$. Entonces para cada entero no negativo n definamos

$$V_n = V_n(x_1, \dots, x_n) = \lim_{N \rightarrow \infty} V_n^N(x_1, \dots, x_n), \quad (2.30)$$

el cual existe μ^n -c.t.p. por el Lema 2.5.

Ahora a estas funciones V_n^N les tomaremos el límite cuando $N \rightarrow \infty$ en el Teorema 2.2 para poder expresar una cota mínima para la función de Lagrange, $\mathcal{L}(\psi)$, dada cualquier regla de paro y la cual puede ser alcanzada bajo ciertas condiciones.

Teorema 2.3. Supongamos que $P(H \geq N) > 0$ para toda $N \geq 1$ y que las condiciones del Corolario 2.2 o el Corolario 2.3 se satisfacen.

Entonces para toda regla de paro ψ

$$\mathcal{L}(\psi) \geq V_0 = p_0 l_0 + (1 - p_0) \min\{l_0, c + J_0 V_1\}. \quad (2.31)$$

Existe igualdad en la desigualdad (2.31) si y sólo si

$$\psi_n \simeq I_{\{l_n \leq c f^n + J_n V_{n+1}\}} \quad \mu^n\text{-c.t.p. en } T_n^\psi \quad (2.32)$$

para toda $n = 0, 1, 2, \dots$

Demostración. Sea $N \geq 1$ cualquier entero no negativo y sea ψ cualquier regla de paro. Es fácil ver, por inducción hacia atrás sobre $k = N, N-1, \dots, 1$, que para toda $1 \leq k \leq N$ se tiene la siguiente desigualdad

$$B_k^\psi(V_k^N) \geq B_{k-1}^\psi(V_{k-1}^N), \quad (2.33)$$

con B_k^ψ definida en (2.19), y V_k^N en (2.23) y (2.24).

Sea ahora k cualquier entero entre 1 y N . Como $B_N^\psi(V_N^N) = B_N^\psi(l_N) = \mathcal{L}_N(\psi)$ y $B_0^\psi(V_0^N) = V_0^N$ es fácil ver a partir de (2.33) que

$$\mathcal{L}_N(\psi) \geq B_k^\psi(V_k^N) \geq B_{k-1}^\psi(V_{k-1}^N) \geq V_0^N \quad (2.34)$$

Como $0 \leq V_k^N \leq l_k$ para toda k , donde $\int l_k d\mu^k \leq \min\{\lambda, 1\} < \infty$ y en virtud de (2.30), $V_k^N \rightarrow V_k$, como $N \rightarrow \infty$, μ^k -c.t.p., se sigue entonces del Teorema de Convergencia Dominada que $B_k^\psi(V_k^N) \rightarrow B_k^\psi(V_k)$, como $N \rightarrow \infty$. Por lo tanto, de (2.34) se sigue que

$$\mathcal{L}(\psi) \geq B_k^\psi(V_k) \geq B_{k-1}^\psi(V_{k-1}) \geq V_0, \quad (2.35)$$

y de (2.24) que

$$V_{k-1} = D_{k-1}V_k \quad (2.36)$$

μ^{k-1} -c.t.p..

La primera parte del Teorema 2.3 está demostrado: se sigue de (2.35) que

$$\mathcal{L}(\psi) \geq V_0,$$

cualquiera que sea la regla de paro ψ .

Supongamos ahora que ψ satisface (2.32) para toda $n = 0, 1, \dots$. Debido a (2.36), se sigue del Lema 2.3 que $B_k^\psi(V_k) = B_{k-1}^\psi(D_{k-1}V_k) = B_{k-1}^\psi(V_{k-1})$ para toda $k = 1, 2, \dots$. En particular, $B_k^\psi(V_k) = B_0^\psi(V_0) = V_0$ (cf. (2.35)), i.e.

$$\sum_{n=0}^{k-1} \int t_n^\psi h_n^\psi (cnf^n + l_n) d\mu^n + P(H \geq k) \int t_k^\psi (ckf^k + V_k) d\mu^k = V_0 \quad (2.37)$$

para toda $k = 1, 2, \dots$. Se sigue ahora de (2.37) que

$$P(H \geq k) \int t_k^\psi ckf^k d\mu^k \leq V_0 (\leq \min\{\lambda, 1\} < \infty),$$

donde la parte izquierda es equivalente a

$$ckP(H \geq k)P(\tau_\psi \geq k) = ckP(\tau_\psi \wedge H \geq k)$$

lo que implica

$$P(\tau_\psi \wedge H \geq k) \leq \frac{\min\{\lambda, 1\}}{ck} \rightarrow 0, \quad k \rightarrow \infty.$$

Así, $P(\tau_\psi \wedge H < \infty) = 1$.

Usando de nueva cuenta (2.37), vemos ahora que

$$\sum_{n=0}^{k-1} \int t_n^\psi h_n^\psi (cnf^n + l_n) d\mu^n \leq V_0$$

para toda $k = 1, 2, \dots$, lo que implica

$$\sum_{n=0}^{\infty} \int t_n^\psi h_n^\psi (cnf^n + l_n) d\mu^n = \mathcal{L}(\psi) \leq V_0.$$

Tomando en cuenta que, por la primera parte del Teorema, $\mathcal{L}(\psi) \geq V_0$, tenemos ahora que $\mathcal{L}(\psi) = V_0$. De esta manera la suficiencia de (2.32) está probada.

La necesidad fácilmente se sigue del mismo Lema 2.3. Si $\mathcal{L}(\psi) = V_0$ para algún ψ , entonces, obviamente, todas son igualdades en (2.35), en particular,

$$B_k^\psi(V_k) = B_{k-1}^\psi(V_{k-1}) = B_{k-1}^\psi(D_{k-1}V_k)$$

(la última ecuación debido a (2.36)). Ahora, esto se sigue del Lema 2.3 que (2.32) se satisface para $n = k - 1$ μ^n -c.t.p. en T_n^ψ . Como $k \geq 1$ es arbitrario, la necesidad de (2.32) se sigue. $\square$

Ahora notemos que si se tiene $P(H \geq N) = 0$ para algún natural N , es fácil ver de (2.11) que no hay más de N sumandos en la expresión (2.15) para $\mathcal{L}(\psi)$, cualquiera que sea ψ . Entonces podemos hacer $N = \max\{n \geq 0 : P(H \geq n) > 0\}$, con lo que tenemos $\mathcal{L}(\psi) = \mathcal{L}_N(\psi)$ (la función Lagrange truncada al nivel N , definida en (2.18)). De esta manera podemos emplear el Teorema 2.2 para minimizar $\mathcal{L}_N(\psi)$ sobre todas las reglas de paro ψ y suponiendo desde luego que $N \geq 1$, de otra manera el problema se vuelve trivial. Se sigue del Teorema 2.3 que, bajo esas condiciones,

$$\inf_{\psi} \mathcal{L}(\psi) = V_0.$$

Así, el Teorema 2.3 caracteriza la estructura de todas las reglas de paro las cuales minimizan $\mathcal{L}(\psi)$.

Enseguida podemos aplicar el Teorema 2.3 al enfoque bayesiano, considerando justamente $P = \pi P_0 + (1 - \pi)P_1$, con $0 < \pi < 1$. Las condiciones de ahí se satisfacen precisamente en virtud del Corolario 2.3, para cualquiera que sea el proceso de observaciones $X_1, X_2, \dots, X_n, \dots$. Suponiendo entonces que $P(H \geq N) > 0$, con $N \geq 1$, y con ayuda del Teorema 2.1, obtenemos de manera inmediata a partir del Teorema 2.3 una caracterización de todas las pruebas secuenciales bayesianas.

Supongamos que $0 < \pi < 1$ es dada. Hagamos $\pi_0 = \pi$ y $\pi_1 = (1 - \pi)$. Tomemos a $f^n = \pi_0 f_0^n + \pi_1 f_1^n$. Entonces por (1.8) vemos que

$$\begin{aligned}
\mathcal{N}(\psi) = E(\tau_\psi \wedge H) &= \sum_{n=0}^{\infty} n E t_n^\psi h_n^\psi \\
&= \sum_{n=0}^{\infty} \int n t_n^\psi h_n^\psi f^n d\mu^n \\
&= \sum_{n=0}^{\infty} \int n t_n^\psi h_n^\psi (\pi f_0^n + (1-\pi) f_1^n) d\mu^n \\
&= \pi \sum_{n=0}^{\infty} \int n t_n^\psi h_n^\psi f_0^n d\mu^n + (1-\pi) \sum_{n=0}^{\infty} \int n t_n^\psi h_n^\psi f_1^n d\mu^n \\
&= \pi_0 E_0(\tau_\psi \wedge H) + \pi_1 E_1(\tau_\psi \wedge H).
\end{aligned}$$

Por lo tanto, combinando los resultados y las demostraciones del Lema 2.1 y los teoremas 2.1 y 2.3 tenemos el siguiente teorema sobre la estructura de las pruebas secuenciales óptimas en el problema original Kiefer-Weiss modificado y que nos será fundamental para los resultados siguientes. El resultado siguiente es una contribución original ya publicada en Novikov & Palacios (2020), que se construye de otros resultados previos a partir del Teorema 2.1.

Teorema 2.4 (Teorema KWMH). *Supongamos que se cumplen las condiciones del Teorema 2.3.*

Sea (ψ, ϕ) , cualquier prueba secuencial, tal que la regla ϕ cumple

$$\phi_n \simeq I_{\{\lambda f_0^n \leq f_1^n\}} \quad \mu^n\text{-c.t.p. en } S_n^\psi \quad (2.38)$$

para toda $n = 0, 1, 2, \dots$, y ψ satisface

$$\psi_n \simeq I_{\{\min\{\lambda f_0^n, f_1^n\} \leq c f^n + J_n V_{n+1}\}} \quad \mu^n\text{-c.t.p. en } T_n^\psi \quad (2.39)$$

para toda $n = 0, 1, \dots$. Entonces para toda prueba secuencial (ψ', ϕ') , tal que

$$\alpha(\psi', \phi') \leq \alpha(\psi, \phi) \quad \text{y} \quad \beta(\psi', \phi') \leq \beta(\psi, \phi) \quad (2.40)$$

se tiene

$$E(\tau_\psi \wedge H) \leq E(\tau_{\psi'} \wedge H). \quad (2.41)$$

La desigualdad en (2.41) es estricta, si al menos una de las desigualdades en (2.40) es estricta.

Si existen igualdades en todas las desigualdades en (2.40) y (2.41) entonces ϕ' también satisface (2.38) (con ϕ' y ψ' en lugar de ϕ y ψ , respectivamente) para toda $n = 0, 1, 2, \dots$, y ψ' satisface (2.39) para toda $n = 0, 1, \dots$, (con ψ' en lugar de ψ).

Es preciso recordar que c , λ y la distribución del horizonte H están involucradas en la construcción de las funciones V_n dadas en el Teorema 2.4 (ver (2.30)). Por lo tanto, dada cualquier prueba secuencial $(\psi, \phi) \in \Delta$, si c_1 y c_2 son dos costos asociados a dicha prueba y λ_1 y λ_2 son dos pesos asociados a la probabilidad del error tipo I ($\alpha(\psi, \phi)$), los cuales

fijamos de antemano, entonces la optimalidad de $\mathcal{L}(\psi, \phi; c_1, \lambda_1)$ y $\mathcal{L}(\psi, \phi; c_2, \lambda_2)$ no tienen por qué ser la misma. Algo parecido sucede al suponer el horizonte aleatorio, H , bajo una distribución F_1 o con una distribución F_2 .

El resto de esta investigación se refiere a aplicar el resultado del Teorema 2.4 a varios modelos estadísticos comúnmente usados en el análisis secuencial, muchos de ellos basados en observaciones i.i.d. o bien sólo independientes. Sin embargo, los resultados de este capítulo, en particular de esta sección, son aplicables a procesos estocásticos más generales que sólo para observaciones independientes. Por ejemplo, no es difícil ver que el Teorema 4.4. en Novikov (2009), puede ser extendido a pruebas de hipótesis para procesos de Markov, cuando el horizonte aleatorio está presente. En particular, los resultados de la Sección 4.5 en Novikov (2009) pueden ser fácilmente extendidos a este modelo de horizonte aleatorio.

Capítulo 3

Optimalidad para Observaciones Independientes

Es esta sección caracterizaremos las pruebas secuenciales óptimas basadas en observaciones independientes, que no necesariamente son idénticamente distribuidas, bajo la presencia de un horizonte aleatorio. Asumamos que las densidades de las observaciones vienen dadas por

$$f_0^n(x_1, \dots, x_n) = \prod_{i=1}^n f_{0,i}(x_i) \text{ y } f_1^n(x_1, \dots, x_n) = \prod_{i=1}^n f_{1,i}(x_i), \quad (3.1)$$

bajo H_0 y H_1 , respectivamente, y que

$$f^n(x_1, \dots, x_n) = \prod_{i=1}^n f_i(x_i) \quad (3.2)$$

es usado como criterio de minimización del tamaño muestral promedio $E(\tau_\psi \wedge H)$.

La optimalidad dada por Wald-Wolfowitz (bien conocida para observaciones i.i.d., ver [Burkholder y Wijsman \(1963\)](#); [Lorden \(1980\)](#); [Wald y Wolfowitz \(1948\)](#)) cuando el número muestral es minimizado bajo ambas hipótesis, es ampliamente deseable para aplicaciones y muy atractivo desde el punto de vista teórico, pero parece ser demasiado fuerte para aplicarlo a modelos más generales, incluso para los más simples que consideramos en esta sección.

3.1. El problema Kiefer-Weiss bajo horizonte aleatorio

En una amplia clase de pruebas de hipótesis para observaciones independientes, las pruebas óptimas derivadas del Teorema [2.4](#) toman una forma mucho más simple. En esta sección veremos que están basadas en “razones de probabilidad” para las distribuciones involucradas.

Supongamos que para $n = 1, 2, \dots$

$$\{f^n = 0\} \subset \{f_0^n = 0\} \cup \{f_1^n = 0\}. \quad (3.3)$$

En cierto sentido lo anterior significa que la distribución correspondiente a f se “encuentra entre” las correspondientes funciones de probabilidad f_0 y f_1 .

Por ejemplo, supongamos que tomamos observaciones independientes provenientes de una distribución uniforme

$$f_\theta(x) = \begin{cases} 1, & \text{si } x \in [\theta, \theta + 1], \\ 0, & \text{en otro caso,} \end{cases} \quad (3.4)$$

y deseamos probar $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$, $\theta_1 > \theta_0$ con un promedio del tamaño mínimo muestral dado por $E\theta\tau_\psi$. Observemos que si $\theta \in [\theta_0, \theta_1]$ entonces se cumple (3.3).

Por lo tanto, si suponemos que la condición (3.3) se satisface, entonces la función de Lagrange $\mathcal{L}(\psi)$ dada en (2.15), puede ser expresada en términos de su razón de probabilidad definida para $n = 1, 2, \dots$ como

$$z_n^0 = z_n^0(x_1, \dots, x_n) = \prod_{i=1}^n \frac{f_{0,i}(x_i)}{f_i(x_i)} \quad \text{y} \quad z_n^1 = z_n^1(x_1, \dots, x_n) = \prod_{i=1}^n \frac{f_{1,i}(x_i)}{f_i(x_i)},$$

y $z_0^0 = z_0^1 \equiv 1$.

De hecho y debido a (3.3) tenemos

$$\begin{aligned} \mathcal{L}(\psi) &= \sum_{n=0}^{\infty} \int t_n^\psi h_n^\psi (cn f^n + \min\{\lambda f_0^n, f_1^n\}) I_{\{f^n > 0\}} d\mu_n \\ &= \sum_{n=0}^{\infty} E t_n^\psi h_n^\psi (cn + \min\{\lambda z_n^0, z_n^1\}). \end{aligned} \quad (3.5)$$

Es preciso indicar que de ahora en adelante, supondremos que cualquier esperanza correspondiente a una función de densidad $f(x)$

$$Eg(X) = \int g(x) f(x) d\mu(x),$$

se entenderá como

$$Eg(X) = \int g(x) f(x) I_{\{f(x) > 0\}} d\mu(x),$$

por lo que no tendremos problemas por preocuparnos por definir $g(x)$ en $\{f(x) = 0\}$. Por lo tanto, no hay problema con la definición de z_n^0 y z_n^1 en (3.5) cuando $f^n = 0$.

En vista de (3.5) parece muy natural que en este caso las reglas de paro óptimas del Teorema 2.4 son funciones sólo de z_n^0 y z_n^1 . Demostremos lo anterior y determinemos la correspondiente estructura óptima de las pruebas en términos de z_n^0 y z_n^1 .

Definamos para $n = 0, 1, 2, \dots$ las funciones $\rho_n^N = \rho_n^N(z^0, z^1)$, $z^0, z^1 > 0$ de la manera siguiente, iniciando con

$$\rho_N^N(z^0, z^1) = g_\lambda(z^0, z^1) \equiv \min\{\lambda z^0, z^1\}, \quad (3.6)$$

y definimos de manera recursiva para $n = N - 1, N - 2, \dots, 1, 0$

$$\rho_n^N(z^0, z^1) = g_\lambda(z^0, z^1)p_n + (1 - p_n) \min \{g_\lambda(z^0, z^1), c + I_n \rho_{n+1}^N(z^0, z^1)\}, \quad (3.7)$$

donde

$$I_n v = I_n v(z^0, z^1) = E v \left(z^0 \frac{f_{0,n+1}(X_{n+1})}{f_{n+1}(X_{n+1})}, z^1 \frac{f_{1,n+1}(X_{n+1})}{f_{n+1}(X_{n+1})} \right), \quad (3.8)$$

para toda función medible no negativa $v = v(z^0, z^1)$.

Lema 3.1. *Bajo la condición (3.3), se tiene para toda $N \geq 1$ y $n = 0, 1, 2, \dots, N$ que*

$$V_n^N = f^n \rho_n^N(z_n^0, z_n^1).$$

Ver demostración en Apéndice. [A.6](#)

Lema 3.2. *Para toda $N \geq 1$ y $n = 0, 1, 2, \dots, N$*

$$\rho_n^N(z^0, z^1) \geq \rho_n^{N+1}(z^0, z^1)$$

para toda $z^0 \geq 0, z^1 \geq 0$.

Ver demostración en Apéndice. [A.7](#)

Por el Lema 3.2, y para toda $n = 0, 1, 2, \dots$ existe

$$\rho_n(z^0, z^1) = \lim_{N \rightarrow \infty} \rho_n^N(z^0, z^1), \quad z^0 \geq 0, z^1 \geq 0.$$

Tomando el límite, como $N \rightarrow \infty$, en (3.7), tenemos que

$$\rho_n(z^0, z^1) = g_\lambda(z^0, z^1)p_n + (1 - p_n) \min \{g_\lambda(z^0, z^1), c + I_n \rho_{n+1}(z^0, z^1)\} \quad (3.9)$$

para toda $n = 0, 1, 2, \dots$

Se sigue ahora del Lema 3.1 que las pruebas secuenciales óptimas en el Teorema 2.4 satisfacen

$$\phi_n \simeq I_{\{\lambda z_n^0 \leq z_n^1\}} \quad \mu^n\text{-c.t.p. en } S_n^\psi \cap \{f^n > 0\}, \quad (3.10)$$

y

$$\psi_n \simeq I_{\{g_\lambda(z_n^0, z_n^1) \leq c + I_n \rho_{n+1}(z_n^0, z_n^1)\}} \quad \mu^n\text{-c.t.p. en } T_n \cap \{f^n > 0\} \quad (3.11)$$

para toda $n = 0, 1, 2, \dots$

Si $f^n = 0$, entonces la prueba óptima en el Teorema 2.4 sugiere seguir tomando observaciones ($\psi^n = 0$), excepto para los casos donde el paro es opcional; i.e. si $f_0^n = 0$ (en cuyo caso se tiene que rechazar H_0 si viene el paro) o $f_1^n = 0$ (aceptar H_0 al detenerse el experimento).

A partir de (3.9) estamos en condiciones de analizar un caso estacionario particular de observaciones independientes.

Podemos hacer en (3.9) que las observaciones sean i.i.d. y además la “tasa de fallos” p_n debida al horizonte aleatorio H sea constante. Por lo tanto y para el resto de esta

sección, supondremos que las observaciones $X_1, X_2, \dots$ son independientes e idénticamente distribuidas, y $p_n = p$ para toda $n = 0, 1, 2, \dots$.

Entonces veamos qué forma tienen las pruebas óptimas (3.10)-(3.11) en este caso en particular.

Para $f_{0,i} = f_0$, $f_{1,i} = f_1$ y $f_i = f$, para toda $i = 1, 2, \dots$, el operador dado en (3.8) toma la forma

$$Iv = Iv(z^0, z^1) = Ev \left(z^0 \frac{f_0(X_1)}{f(X_1)}, z^1 \frac{f_1(X_1)}{f(X_1)} \right) \quad (3.12)$$

para toda función medible no negativa $v = v(z^0, z^1)$.

De las funciones (3.6)-(3.7) podemos ver que bajo las condiciones de esta sección se cumple

$$\rho_n^N(z^0, z^1) = \rho_{n+1}^{N+1}(z^0, z^1), \quad z^0, z^1 \geq 0, \quad (3.13)$$

para toda $n \leq N$ y toda N .

Tomando el límite en (3.13), como $N \rightarrow \infty$, tenemos

$$\rho_n(z^0, z^1) = \rho_{n+1}(z^0, z^1), \quad z^0, z^1 \geq 0,$$

para toda $n = 0, 1, 2, \dots$. Si denotamos a $\rho(z^0, z^1) = \rho_n(z^0, z^1)$, $z^0, z^1 \geq 0$, obtenemos de (3.9)

$$\rho(z^0, z^1) = g_\lambda(z^0, z^1)p + (1-p) \min\{g_\lambda(z^0, z^1), c + I\rho(z^0, z^1)\} \quad (3.14)$$

para toda $z^0, z^1 \geq 0$.

Es así que las pruebas secuenciales del Teorema 2.4 (ver (3.10)-(3.11)) deben satisfacer ahora

$$\phi_n \simeq I_{\{\lambda z_n^0 \leq z_n^1\}} \quad \mu^n\text{-c.t.p. en } S_n^\psi \cap \{f^n > 0\}, \quad (3.15)$$

y

$$\psi_n \simeq I_{\{g_\lambda(z_n^0, z_n^1) \leq c + I\rho(z_n^0, z_n^1)\}} \quad \mu^n\text{-c.t.p. en } T_n^\psi \cap \{f^n > 0\} \quad (3.16)$$

para toda $n = 0, 1, 2, \dots$.

Como era de esperarse, cuando $p = 0$ corresponde a un horizonte infinito $H \equiv \infty$ y nuestras pruebas secuenciales (3.15)-(3.16) son versiones aleatorizadas (extendidas, en los términos de [Burkholder y Wijsman \(1963\)](#)) de la prueba óptima (3) de [Lorden \(1980\)](#).

Para finalizar esta sección es conveniente ilustrar los resultados obtenidos en esta investigación para la construcción de pruebas secuenciales óptimas bajo horizonte en el problema Kiefer-Weiss.

Supongamos que las observaciones de un proceso tienen todas distribución uniforme (3.4) y deseamos realizar la siguiente prueba de hipótesis $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$, con $\theta_1 > \theta_0$. Nuestro propósito será entonces minimizar el máximo tamaño promedio muestral

$$\sup_{\theta \in [\theta_0, \theta_1]} E_\theta \tau_\psi.$$

Este es el problema Kiefer-Weiss original, pero ahora bajo horizonte Aleatorio. Además, supongamos que el horizonte aleatorio H tiene una tasa de fallos constante, i.e.

$$P(H = n) = p(1-p)^n, \quad \forall n = 0, 1, \dots \quad (3.17)$$

(distribución geométrica con parámetro p).

Para un parámetro θ fijo, con $\theta_0 < \theta < \theta_1$, construyamos la prueba óptima minimizando $E_\theta(\tau_\psi \wedge H)$ definida por (3.15) – (3.16).

Para empezar y por cuestiones prácticas hagamos $\lambda = 1$, entonces $0 < c < \theta_1 - \theta_0$ de tal manera que podamos iniciar el análisis inductivo hacia atrás (más adelante se ve la razón en el Corolario 3.2). Enseguida, necesitamos la forma del operador I en (3.16). Sea $v(z^0, z^1)$ una función medible no negativa, tal que $v(0, z^1) = v(z^0, 0) = 0$ para todos los valores no negativos z^0 y z^1 . Es así que por definición tenemos

$$Iv(z^0, z^1) = \int_{\theta}^{\theta+1} v\left(z^0 \frac{f_{\theta_0}}{f_{\theta}}, z^1 \frac{f_{\theta_1}}{f_{\theta}}\right) f_{\theta}(x) dx = v(z^0, z^1) x \Big|_{\theta_1}^{\theta_0+1} = v(z^0, z^1)(1 + \theta_0 - \theta_1),$$

por que $f^n > 0$ en $[\theta_1, \theta_0 + 1]$. Por lo tanto, (3.14) se convierte en

$$\rho(z^0, z^1) = g_1(z^0, z^1)p + (1 - p) \min\{g_1(z^0, z^1), c + \rho(z^0, z^1)(1 + \theta_0 - \theta_1)\} \quad (3.18)$$

para todas $z^0, z^1 \geq 0$ y

$$\psi_n \simeq I_{\{g_1(z_n^0, z_n^1) \leq c + \rho(z_n^0, z_n^1)(1 + \theta_0 - \theta_1)\}} \quad (3.19)$$

para toda $n = 0, 1, 2, \dots$

Por la misma definición, z_n^0 y z_n^1 toman sólo valores 0 ó 1. Podemos verificar fácilmente que si uno de estos es 0, entonces la regla de paro óptima (3.19) es igual a 1, y se deberá detener el procedimiento. Ahora demostremos que si $z_n^0 = 1$ y $z_n^1 = 1$, entonces no se detendrá el proceso, al menos cuando c es lo suficientemente pequeño. Debido a (3.18) $\rho(1, 1) \leq 1$, entonces evaluando (3.19) tenemos

$$c + \rho(1, 1)(1 + \theta_0 - \theta_1) \leq c + (1 + \theta_0 - \theta_1) < 1$$

si $c < (\theta_1 - \theta_0)$. Así, la regla de paro óptimo (3.19) tiene que continuar el procedimiento cuando $z_n^0 = 1$ y $z_n^1 = 1$.

Vemos que la regla de paro óptima en (3.19) tiene que detener el procedimiento al primer tiempo n cuando $z_n^0 = 0$ o $z_n^1 = 0$ y deberá de continuar en otro caso. La prueba también debe parar en la etapa n cuando $H = n$, inclusive cuando ambas $z_n^0 = 1$ y $z_n^1 = 1$, por ser un paro independiente de las observaciones. La regla decisión se aplica en este caso (3.15), así deberíamos rechazar H_0 (porque $\lambda = 1$).

Caracterizemos el procedimiento óptimo que justamente hemos determinado. Primero hagámoslo para el tamaño promedio muestral, pero observemos que $\theta_1 - \theta_0 > 0$ luego $1 - \theta_0 + \theta_1 < 1$ y también $1 - p < 1$. entonces:

$$\begin{aligned}
E_\theta(\tau_\psi \wedge H) &= \sum_{n=1}^{\infty} P_\theta(\tau_\psi \geq n, H \geq n) = \sum_{n=1}^{\infty} P_\theta(\tau_\psi \geq n)P(H \geq n) \\
&= \sum_{n=1}^{\infty} (1 + \theta_0 - \theta_1)^{n-1} (1-p)^n \\
&= \lim_{k \rightarrow \infty} \sum_{n=1}^k (1 + \theta_0 - \theta_1)^{n-1} (1-p)^n \\
&= \lim_{k \rightarrow \infty} \frac{(1 + \theta_0 - \theta_1)^{k+1} (1-p)^{k+2} - (1-p)}{(1 + \theta_0 - \theta_1)(1-p) - 1} \\
&= \frac{p-1}{1-p + (\theta_0 - \theta_1)(1-p) - 1} \\
&= \frac{1-p}{p + (\theta_1 - \theta_0)(1-p)}.
\end{aligned}$$

Y ahora para las probabilidades de los errores tipo I y II, tenemos

$$\alpha(\psi, \phi) = \sum_{n=0}^{\infty} E_{\theta_0} \phi_n t_n \psi_n P(H > n) + \sum_{n=0}^{\infty} E_{\theta_0} \phi_n t_n P(H = n) \quad (3.20)$$

$$= \sum_{n=0}^{\infty} E_{\theta_0} I_{\{f_{\theta_1}^n \geq f_{\theta_0}^n\}} I_{\{X_i \in [\theta_1, 1+\theta_0], i=1,2,\dots,n-1\}} I_{\{X_n \in [\theta_0, \theta_1]\}} P(H > n) \quad (3.21)$$

$$+ \sum_{n=0}^{\infty} E_{\theta_0} I_{\{f_{\theta_1}^n \geq f_{\theta_0}^n\}} I_{\{X_i \in [\theta_1, 1+\theta_0], i=1,2,\dots,n-1\}} P(H = n) \quad (3.22)$$

$$= \sum_{n=0}^{\infty} (1 + \theta_0 - \theta_1)^n p (1-p)^n = \frac{p}{1 - (1 + \theta_0 - \theta_1)(1-p)} \quad (3.23)$$

$$= \frac{p}{p + (\theta_1 - \theta_0)(1-p)} \quad (3.24)$$

$$\beta(\psi, \phi) = \sum_{n=0}^{\infty} E_{\theta_1} (1 - \phi_n) t_n \psi_n P(H > n) + \sum_{n=0}^{\infty} E_{\theta_1} (1 - \phi_n) t_n P(H = n) \quad (3.25)$$

$$= \sum_{n=0}^{\infty} E_{\theta_1} I_{\{f_{\theta_1}^n < f_{\theta_0}^n\}} I_{\{X_i \in [\theta_1, 1+\theta_0], i=1,2,\dots,n-1\}} I_{\{X_n \in [\theta_0, \theta_1]\}} P(H > n) \quad (3.26)$$

$$+ \sum_{n=0}^{\infty} E_{\theta_1} I_{\{f_{\theta_1}^n < f_{\theta_0}^n\}} I_{\{X_i \in [\theta_1, 1+\theta_0], i=1,2,\dots,n-1\}} P(H = n) = 0. \quad (3.27)$$

En conclusión, tenemos que debido a la presencia del horizonte aleatorio, los resultados son:

$$E_\theta(\tau_\psi \wedge H) = \frac{1-p}{p + (\theta_1 - \theta_0)(1-p)}, \quad (3.28)$$

$$\alpha(\psi, \phi) = \frac{p}{p + (\theta_1 - \theta_0)(1-p)}, \quad (3.29)$$

$$\beta(\psi, \phi) = 0. \quad (3.30)$$

Pero para un horizonte aleatorio infinito ($H \equiv \infty$), equivalente a tomar $p = 0$, tenemos

$$\begin{aligned} E_\theta(\tau_\psi) &= \frac{1}{\theta_1 - \theta_0}, \\ \alpha(\psi, \phi) &= 0, \quad \beta(\psi, \phi) = 0. \end{aligned} \quad (3.31)$$

En este ejemplo simple, podemos fácilmente estimar el efecto de un horizonte aleatorio en el desempeño de la prueba óptima. Si no hay un horizonte finito, ambas probabilidades de los errores son iguales a 0, y el tamaño muestral es el recíproco de $\theta_1 - \theta_0$. Ahora bien, cuando la distribución del horizonte es geométrica, $\alpha(\psi, \phi)$ aumenta a (3.29), y el número muestral promedio decrece como en (3.28). Por lo tanto, el decaimiento en el tamaño muestral promedio, debido al efecto del horizonte, es

$$\frac{1}{\theta_1 - \theta_0} - \frac{1 - p}{p + (\theta_1 - \theta_0)(1 - p)} = \frac{p}{(\theta_1 - \theta_0)(p + (\theta_1 - \theta_0)(1 - p))}.$$

Es así que, si la probabilidad del error tipo I aumenta a α , debido a un horizonte aleatorio, el número muestral promedio se reducirá en $\alpha/(\theta_1 - \theta_0)$, y viceversa : teniendo información confiable sobre el horizonte aleatorio, podríamos reducir el número muestral promedio por algún valor r , pero es pertinente aclarar que esto causará el aumento de la probabilidad del error a $r(\theta_1 - \theta_0)$.

Lo último de esta sección por hacer, es notar que la prueba que obtuvimos en este ejemplo en efecto resuelve el problema original Kiefer-Weiss para este modelo uniforme bajo un horizonte aleatorio geométrico.

Sean α y β definidos por (3.29) y (3.30), respectivamente. Además supongamos que hay otra prueba (ψ', ϕ') tal que

$$\alpha(\psi', \phi') \leq \alpha \quad \text{y} \quad \beta(\psi', \phi') = 0.$$

Entonces, y debido a la optimalidad de (ψ, ϕ) para θ tenemos

$$\sup_{\theta \in [\theta_0, \theta_1]} E_\theta(\tau_{\psi'} \wedge H) \geq E_\theta(\tau_{\psi'} \wedge H) \geq E_\theta(\tau_\psi \wedge H) = \sup_{\theta \in [\theta_0, \theta_1]} E_\theta(\tau_\psi \wedge H),$$

entonces la prueba (ψ, ϕ) es óptima en el problema Kiefer-Weiss en presencia del horizonte geométrico.

Además, si el horizonte es infinito, la misma prueba (ψ, ϕ) es óptima para cualquier $\theta \in [\theta_0, \theta_1]$ para $\alpha = 0$ y $\beta = 0$, por construcción. Por lo tanto, es óptima en el problema Kiefer-Weiss con $\alpha = 0$ y $\beta = 0$ y también bajo el horizonte infinito.

3.2. Minimización del promedio muestral bajo una hipótesis

Supongamos que, bajo la hipótesis H_i , las densidades (con respecto a μ en $\mathfrak{X}$) de las observaciones independientes $X_1, X_2, X_3, \dots$, son respectivamente, $f_{i,1}, f_{i,2}, f_{i,3} \dots$, $i = 0, 1$. El enfoque clásico de observaciones independientes e idénticamente distribuidas es

un caso particular de ésta, cuando $f_{i,j} = f_i$ para toda j e $i = 0, 1$. Algunos modelos aplicados involucran la independencia, no así observaciones idénticamente distribuidas, como se describe en [Liu y Blostein \(1992\)](#). De hecho los autores afirmaron que la prueba que ellos construyeron minimiza el tamaño muestral promedio no solo para H_0 sino también para H_1 (optimalidad de [Wald y Wolfowitz \(1948\)](#)), pero, debido a [Schmitz \(1995\)](#), no es cierto en general.

En esta sección, deseamos ver cuándo la optimalidad toma lugar para alguna de las hipótesis, en presencia de un horizonte aleatorio. Queremos construir las pruebas secuenciales óptimas que minimizan el número muestral promedio bajo una de las hipótesis, digamos bajo H_0 . Esto significa que nuestro criterio de minimización será $\mathcal{N}(\psi) = E_0(\tau_\psi \wedge H)$.

Todos los resultados previos se aplican inmediatamente tomando $P = P_0$, lo cual nos indica que debemos utilizar a f_i^n y f^n en esta sección, respectivamente, como

$$f_i^n = f_i^n(x_1, \dots, x_n) = \prod_{k=1}^n f_{i,k}(x_k), \quad i = 0, 1 \quad \text{y} \quad f^n = f^n(x_1, \dots, x_n) = \prod_{k=1}^n f_{0,k}(x_k). \quad (3.32)$$

Por razones técnicas, supondremos a lo largo de esta sección que la distribución de las observaciones X_n bajo H_0 y H_1 son equivalentes (mutuamente absolutamente continuas), es decir,

$$P_i(f_{1-i,n}(X_n) = 0) = 0 \quad \text{para } i = 0, 1, \quad \text{y para toda } n = 1, 2, \dots \quad (3.33)$$

Sin el supuesto (3.33), la caracterización de las pruebas secuenciales óptimas se vuelven algo complicadas, pero aun así es bastante factible sobre la base de los resultados del Capítulo 2

Introduzcamos la siguiente familia de funciones $\rho_n^N : \mathbb{R}^+ \mapsto \mathbb{R}^+$, $0 \leq n \leq N$, para cualquier entero no negativo N , que nos ayudará en el análisis de esta parte de la investigación.

Para cualquier función medible no negativa $v : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, definamos

$$I_nv = I_nv(z) = E_0v \left(z \frac{f_{1,n+1}(X_{n+1})}{f_{0,n+1}(X_{n+1})} \right), \quad z \geq 0, \quad (3.34)$$

para $n = 0, 1, \dots$. Sea $g_\lambda(z)$ definida como $g_\lambda(z) = \min\{\lambda, z\}$ para toda $z \geq 0$. Ahora bien, para cualquier $N \geq 0$, hagamos

$$\rho_N^N(z) = g_\lambda(z), \quad z \geq 0, \quad (3.35)$$

y, recursivamente, para $n = N - 1, \dots, 1, 0$,

$$\rho_n^N(z) = g_\lambda(z)p_n + (1 - p_n) \min\{g_\lambda(z), c + I_n\rho_{n+1}^N(z)\}, \quad (3.36)$$

para toda $z \in \mathbb{R}^+$.

Tengamos en cuenta que todas las funciones ρ_n^N dependen de los números positivos fijos c y λ , i.e. $\rho_n^N(z) = \rho_n^N(z; c, \lambda)$, $z \geq 0$, para toda $N \geq 0$ y $0 \leq n \leq N$. Como antes, p_n se entiende que es la tasa de fallos debida al horizonte aleatorio.

Para una λ dada, listaremos todas las propiedades de las funciones con las que vamos a trabajar. Sea $\mathcal{K} = \mathcal{K}(\lambda)$ la clase de las funciones $v : \mathbb{R}^+ \mapsto \mathbb{R}^+$ tal que

- (a) v es no decreciente, cóncava y continua en $\mathbb{R}^+$,
- (b) $v(0) = 0$, $\lim_{z \rightarrow \infty} v(z) = \lambda$ y
- (c) $v(z) \leq \min\{\lambda, z\}$ para toda $z \in \mathbb{R}^+$.

Lema 3.3.

- (i) Para todo entero no negativo N , $\rho_n^N \in \mathcal{K}$ para toda $0 \leq n \leq N$ y $I_n \rho_{n+1}^N \in \mathcal{K}$ para toda $0 \leq n \leq N-1$.
- (ii) Para todo $z \in \mathbb{R}^+$ $\rho_n^N(z) \geq \rho_n^{N+1}(z)$, cualquiera que sea $N \geq 0$ y $0 \leq n \leq N$.
- (iii) Para toda $n \geq 0$ fija, el límite $\rho_n(z) = \lim_{N \rightarrow \infty} \rho_n^N(z)$ existe para todo $z \geq 0$. Las funciones límite $\rho_n \in \mathcal{K}$, y $I_n \rho_{n+1} \in \mathcal{K}$, para toda $n \geq 0$.
- (iv) Para toda $n \geq 0$

$$\rho_n(z) = g_\lambda(z)p_n + (1 - p_n) \min\{g_\lambda(z), c + I_n \rho_{n+1}(z)\}, \quad (3.37)$$

para todo $z \in \mathbb{R}^+$.

Ver demostración en el apéndice A.8.

Probemos ahora que todos los elementos de las reglas de paro óptimas debido a los Teoremas 2.2 y 2.3, pueden ser expresados en términos de las funciones $\rho_n^N(Z_n)$, donde Z_n es el estadístico de la razón de verosimilitud, definido, para $n = 0, 1, 2, \dots$, como

$$Z_n = Z_n(x_1, x_2, \dots, x_n) = \begin{cases} \frac{f_1^n(x_1, \dots, x_n)}{f_0^n(x_1, \dots, x_n)}, & \text{si } f_0^n(x_1, \dots, x_n) > 0, \\ 0, & \text{en otro caso,} \end{cases}$$

donde f_i^n , $i = 0, 1$, corresponden a observaciones independientes (ver (3.32)), y por definición, $Z_0 = 1$.

Las funciones V_n^N y V_n que definimos en el capítulo anterior (ver (2.23), (2.24) y (2.30), respectivamente), en este caso de observaciones independientes, adquieren fácilmente la siguiente forma.

Lema 3.4. Para toda $0 \leq n \leq N$

$$V_n^N = f_0^n \rho_n^N(Z_n), \quad (3.38)$$

cualquiera que sea $N \geq 0$. Para toda $n \geq 0$,

$$V_n = f_0^n \rho_n(Z_n). \quad (3.39)$$

La demostración de (3.38), se sigue de manera análoga a la prueba del Lema 3.1, mientras que (3.39) se sigue de (3.38) tomando el límite cuando $N \rightarrow \infty$.

Aquí la parte importante del por qué relacionar estas funciones ρ_n^N y V_n^N , es por el hecho de aprovechar las propiedades de las funciones ρ_n^N del Lema 3.3, ya que éstas quedan definidas en todo $\mathbb{R}^+$; en particular podemos aplicar algunas de ellas en los puntos de los

estadísticos de razón de verosimilitud, $Z_n \geq 0$, las cuales usaremos para encontrar, más adelante y de manera única, los valores que se conocen como región de continuación del experimento secuencial.

Una consecuencia inmediata que obtenemos del Lema 3.4, en particular, que la cota inferior del Teorema 2.3 puede ser expresada de una manera equivalente, como $V_0 = \rho_0(1)$. Tomando en cuenta la definición de $\mathcal{L}(\psi)$ en (2.14), y (2.2), tenemos que, bajo las condiciones del Teorema 2.3,

$$\inf_{\psi, \phi} \{c\mathcal{N}(\psi) + \lambda\alpha(\psi, \phi) + \beta(\psi, \phi)\} = \rho_0(1; c, \lambda). \quad (3.40)$$

Un hecho interesante es que esto puede ser generalizado de la manera siguiente.

Lema 3.5. *Bajo las condiciones del Teorema 2.3*

$$\inf_{\psi, \phi} \{c\mathcal{N}(\psi) + \lambda\alpha(\psi, \phi) + z\beta(\psi, \phi)\} = \rho_0(z; c, \lambda) \quad (3.41)$$

para toda $z > 0$, $c > 0$, $\lambda > 0$.

Demostración. Como $z > 0$, tenemos de (3.40) que

$$\begin{aligned} & \inf_{\psi, \phi} \{c\mathcal{N}(\psi) + \lambda\alpha(\psi, \phi) + z\beta(\psi, \phi)\} \\ &= z \inf_{\psi, \phi} \left\{ \frac{c}{z} \mathcal{N}(\psi) + \frac{\lambda}{z} \alpha(\psi, \phi) + \beta(\psi, \phi) \right\} = z\rho_0(1; c/z, \lambda/z), \end{aligned}$$

la afirmación del Lema se sigue fácilmente tomando el límite como $N \rightarrow \infty$ en

$$z\rho_n^N(1; c/z, \lambda/z) = \rho_n^N(z; c, \lambda),$$

lo cual, a su vez, se verifica de manera inmediata a través de inducción hacia atrás. $\square$

Notemos que el ínfimo en (3.41) es aplicado a la familia de funciones lineales, con respecto a (z, c, λ) , de esta manera vemos que la función límite, $\rho_0(z; c, \lambda)$, es una función cóncava de (z, c, λ) . Por lo tanto, aplicando el Teorema 10.1 en Rockafellar (1970) tenemos el siguiente.

Corolario 3.1. *La función $\rho_0(z; c, \lambda)$ es continua, como función de sus tres argumentos (z, c, λ) , en $\{z > 0, c > 0, \lambda > 0\}$.*

Observemos que el Corolario 3.1 se cumple, no sólo para ρ_0 , sino también para toda ρ_n ; justamente aplicando los resultados previos a la sucesión de observaciones $X_{n+1}, X_{n+2}, \dots$ y condicionando la distribución del horizonte H en $H \geq n$, $n \geq 0$.

Podemos ver, a partir del Lema 3.4, que

$$\begin{aligned} J_n V_{n+1} &= \int V_{n+1}(x_1, \dots, x_{n+1}) d\mu(x_{n+1}) \\ &= \int f_0^{n+1} \rho_{n+1}(Z_{n+1}) d\mu(x_{n+1}) \\ &= f_0^n \int \rho_{n+1} \left(Z_n \frac{f_{1,n+1}(x_{n+1})}{f_{0,n+1}(x_{n+1})} \right) f_{0,n+1}(x_{n+1}) d\mu(x_{n+1}) \\ &= f_0^n E_0 \rho_{n+1}(Z_n) = f_0^n I_n \rho_{n+1}(Z_n). \end{aligned}$$

Por lo tanto, tenemos de manera inmediata que la desigualdad que define la regla de paro óptima en el Teorema 2.3, es

$$l_n \leq cf_0^n + J_n V_{n+1}$$

(ver (2.32)), es equivalente a

$$g_\lambda(Z_n) \leq c + I_n \rho_{n+1}(Z_n), \quad (3.42)$$

cuando $f_0^n > 0$.

A partir del siguiente lema se demuestra que los conjuntos de Z_n para el que las desigualdades del tipo (3.42) mantienen una estructura muy simple (del tipo GSPRT, – ver Eisenberg et al. (1976)).

Lema 3.6. *Si $n \geq 0$ es tal que*

$$\lambda > c + I_n \rho_{n+1}(\lambda), \quad (3.43)$$

entonces existen constantes A_n y B_n , $A_n < \lambda < B_n$, tales que

$$A_n = c + I_n \rho_{n+1}(A_n), \quad y \quad \lambda = c + I_n \rho_{n+1}(B_n),$$

y la desigualdad

$$g_\lambda(z) \leq c + I_n \rho_{n+1}(z) \quad (3.44)$$

se satisface si y sólo si $z \notin (A_n, B_n)$.

Si (3.43) no se cumple, entonces (3.44) es válido para todo $z \geq 0$.

La demostración se encuentra en el Apéndice A A.9.

Procedamos ahora con la caracterización de todas las reglas de paro óptimas en el caso de observaciones independientes bajo el modelo de un horizonte aleatorio.

Definamos la variable aleatoria $\mathcal{Z}_n$ como $\mathcal{Z}_n = Z_n(X_1, \dots, X_n)$, y sean los eventos $\mathcal{T}_n^\psi$ y $\mathcal{S}_n^\psi$ como $\mathcal{T}_n^\psi = \{(X_1, \dots, X_n) \in T_n^\psi\}$ y $\mathcal{S}_n^\psi = \{(X_1, \dots, X_n) \in S_n^\psi\}$, respectivamente para $n = 0, 1, \dots$. Sabemos que dado un vector de observaciones $(x_1, \dots, x_n) \in T_n^\psi$ es $\mathfrak{X}^n$ –medible, tal que $t_n^\psi(x_1, \dots, x_n) > 0$, es decir,

$$P(\tau_\psi \geq n | X_1, \dots, X_n) = t_n^\psi > 0.$$

Lo que se traduce a que el experimento no termina por la regla de paro de la prueba secuencial sino por el horizonte aleatorio con probabilidad t_n^ψ en la etapa n , por lo que

$$\mathcal{T}_n^\psi = \{P(\tau_\psi \geq n | X_1, \dots, X_n) > 0\}.$$

Algo análogo obtenemos para $\mathcal{S}_n^\psi$. El vector de observaciones $(x_1, \dots, x_n) \in S_n^\psi$ es $\mathfrak{X}^n$ –medible, tal que $t_n^\psi(x_1, \dots, x_n)h_n^\psi(x_1, \dots, x_n) > 0$ (ver (2.11)). Entonces

$$P(\tau_\psi = n | X_1, \dots, X_n) = t_n^\psi h_n^\psi > 0,$$

que es la probabilidad de que el experimento termine debido a la regla de paro y no antes y ni por el horizonte aleatorio. Por lo tanto,

$$\mathcal{S}_n^\psi = \{P(\tau_\psi = n | X_1, \dots, X_n) > 0\}.$$

De ahora en adelante “c.s.” se entenderá como “casi seguramente”. Si una condición se cumple c.s., entonces existe un conjunto medible A en donde la condición no se satisface y este conjunto es improbable o imposible de ocurrir, i.e. $P(A) = 0$.

Con lo antes visto en esta sección obtenemos y del Teorema 2.3, tenemos el siguiente teorema.

Teorema 3.1. *Supongamos que $P(H \geq n) > 0$ para toda $n \geq 1$ y también se satisface (3.43) para toda n .*

Entonces para toda regla de paro ψ se tiene

$$\mathcal{L}(\psi) \geq \rho_0(1). \quad (3.45)$$

Existe igualdad en (3.45) si y sólo si

$$I_{\{Z_n \notin [A_n, B_n]\}} \leq \psi_n(X_1, \dots, X_n) \leq I_{\{Z_n \in (A_n, B_n)\}} \quad P_0\text{-c.s. en } \mathcal{T}_n^\psi \quad (3.46)$$

para todo entero no negativo n .

Demostración. Las condiciones del Teorema 2.3 se satisfacen, porque, en este caso hemos considerado a $P = P_0$, entonces la condición (i) del Corolario 2.3 se satisface de manera trivial. Ahora (3.45) se sigue de (2.31) en el Teorema 2.3, porque $V_0 = \rho_0(1)$, debido a (3.39). Finalmente (3.46) es equivalente a (2.32) debido al Lema 3.6 y por el hecho de que

$$(x_1, \dots, x_n) \in A \quad \mu^n\text{-c.t.p. en } B \cap \{f_0^n(x_1, \dots, x_n) > 0\}$$

es equivalente a

$$(X_1, \dots, X_n) \in A \quad P_0\text{-c.s. en } \{(X_1, \dots, X_n) \in B\},$$

para cualesquiera que sean $A, B \in \mathfrak{X}^n$. □

Combinando el Lema 2.1 y los teorema 2.1 y 3.1 obtenemos inmediatamente el siguiente teorema.

Teorema 3.2. *Supongamos que $P(H \geq n) > 0$ para toda $n \geq 1$ y que la condición (3.43) se satisface.*

Sea (ψ, ϕ) cualquier prueba secuencial, tal que ψ satisface (3.46) para toda $n \geq 0$, y ϕ satisface

$$\phi_n(X_1, \dots, X_n) \simeq I_{\{Z_n \geq \lambda\}} \quad P_0\text{-c.s. en } \mathcal{S}_n^\psi \quad (3.47)$$

para toda $n \geq 0$. Entonces para toda prueba secuencial (ψ', ϕ') , tal que

$$\alpha(\psi', \phi') \leq \alpha(\psi, \phi) \quad \text{y} \quad \beta(\psi', \phi') \leq \beta(\psi, \phi) \quad (3.48)$$

se cumple

$$E_0(\tau_\psi \wedge H) \leq E_0(\tau_{\psi'} \wedge H). \quad (3.49)$$

La desigualdad en (3.49) es estricta si al menos una de las desigualdades en (3.48) es estricta.

Si hay igualdades en todas las desigualdades de (3.48) y (3.49) entonces

$$I_{\{Z_n \notin [A_n, B_n]\}} \leq \psi'_n(X_1, \dots, X_n) \leq I_{\{Z_n \notin (A_n, B_n)\}} \quad P_0\text{-c.s. en } \mathcal{T}_n^{\psi'} \quad (3.50)$$

para toda $n \geq 0$ y

$$\phi'_n(X_1, \dots, X_n) \simeq I_{\{Z_n \geq \lambda\}} \quad P_0\text{-c.s. en } \mathcal{S}_n^{\psi'} \quad (3.51)$$

para toda $n \geq 0$.

Notemos que las pruebas secuenciales óptimas truncadas, digamos a un nivel $N \geq 0$, admiten una caracterización similar. La diferencia es que ρ_{n+1}^N debe ser usado en lugar de ρ_{n+1} en el Lema 3.6, y consecuentemente, las constantes para la llamada región de continuación serán A_n^N y B_n^N . Finalmente, las reglas de paro en, (3.46) y (3.47), se restringen a una $n \leq N$, mientras que para las reglas de decisión (3.47) y (3.51) a una $n \leq N - 1$, asumiendo desde luego para este caso que $\psi_N = 1$.

Tanto en el Teorema 3.1 como en el Teorema 3.2 hemos supuesto que se satisfacen ciertas condiciones, la primera que $P(H \geq n) > 0$ para toda $n \geq 1$ y que se cumple (3.43). En caso que $P(H \geq n) = 0$ para alguna $n \geq 1$, entonces la prueba secuencial óptima (ψ, ϕ) del Teorema 3.2 será truncada al nivel

$$N = \max\{n \geq 0 : P(H \geq n) > 0\}.$$

Por otro lado, si $P(H \geq n) > 0$ para toda $n \geq 1$, pero (3.43) no se satisface para alguna $n \geq 0$, entonces la prueba secuencial óptima puede ser nuevamente encontrada en la clase de pruebas truncadas, tomando a N precisamente como

$$N = \min\{n \geq 0 : \lambda \leq c + I_n \rho_{n+1}(\lambda)\}.$$

Debido a (3.40), el valor $\rho_0(1) = \rho_0(1; c, \lambda)$ es el mínimo de la función de Lagrange calculada bajo la presencia de H . Denotemos esta cota como $\rho_0^H(1; c, \lambda)$ y sea $\rho_0^*(1; c, \lambda)$ la función de Lagrange calculada bajo este supuesto. Es fácil ver que $\rho_0^*(z; c, \lambda) \leq \rho_0^H(z; c, \lambda)$ para toda $z \geq 0$ y la diferencia $\rho_0^H(1; c, \lambda) - \rho_0^*(1; c, \lambda)$ puede ser considerada como el impacto del horizonte en el problema de la prueba. Además, si tomamos esperanza en (3.40), no es difícil ver que la diferencia converge en distribución a 0 cuando $H \rightarrow \infty$, por lo que la prueba correspondiente $\rho_0^*(1; c, \lambda)$ puede ser considerada una solución asintótica al problema de la prueba cuando $H \rightarrow \infty$.

Finalmente, apliquemos los resultados al caso de observaciones i.i.d. y bajo un horizonte aleatorio con distribución geométrica.

3.3. Observaciones i.i.d

De ahora en adelante, supondremos que el horizonte aleatorio posee distribución geométrica con parámetro p . La caracterización de las pruebas secuenciales óptimas en el caso

de observaciones i.i.d, tiene aún una forma más sencilla. Sea $f_{i,k}(x) = f_i(x)$, para toda $k = 1, 2, \dots$, $i = 0, 1$, es decir, tomando las densidades de las hipótesis

$$f_i^n(x_1, \dots, x_n) = \prod_{k=1}^n f_i(x_k), \quad i = 0, 1, \quad (3.52)$$

y usando $E_0(\tau_\psi \wedge H)$ (o bien $E_1(\tau_\psi \wedge H)$) como criterio para la minimización del número promedio muestral.

Supongamos que las distribuciones hipotetizadas son distintas, i.e

$$\mu(\{x : f_0(x) \neq f_1(x)\}) > 0. \quad (3.53)$$

La condición de equivalencia que necesitamos en este caso es

$$P_i(f_{1-i}(X) = 0) = 0, \quad i = 0, 1, \quad (3.54)$$

A pesar de que el modelo de observaciones i.i.d. (3.52) parece ser mucho más simple, que el estudiado en general de la sección anterior, no hay un debilitamiento aparente, generalmente hablando, en la estructura de las pruebas óptimas en el Teorema 3.2, principalmente debido a la presencia de la variación de la “tasa de fallos” p_n de la distribución del horizonte aleatorio el cual participa en las definiciones de las funciones ρ_n^N . Podemos decir que la única y afortunada excepción, del caso ocurre cuando $p_n = p$, $p \in (0, 1)$, para toda $n = 0, 1, \dots$, lo cual corresponde a la distribución *geométrica* (con parámetro p) del horizonte aleatorio, que anteriormente ya habíamos mencionado. Es fácil ver que en este caso

$$P(H = n) = p(1 - p)^n, \quad n = 0, 1, 2, \dots \quad (3.55)$$

El significado aplicado de esta distribución es bien conocido: si H obedece a (3.55), entonces puede ser interpretado como el tiempo del primer éxito en una serie de ensayos independientes de Bernoulli con probabilidad p de éxito en cada uno. De esta manera, corresponde al caso cuando la intervención, debido al horizonte, ocurre para cualquier etapa del experimento secuencial con la misma probabilidad p , independientemente de que ocurra o no en otras etapas.

Recordemos que la tasa de fallos es $p_n = \frac{P(H=n)}{P(H \geq n)}$ para toda $n \geq 0$, de donde

$$\begin{aligned} P(H \geq n) &= \sum_{k=n}^{\infty} P(H = k) = p(1 - p)^n + p(1 - p)^{n+1} + p(1 - p)^{n+2} + \dots \\ &= p(1 - p)^n (1 + (1 - p) + (1 - p)^2 + \dots) \\ &= p(1 - p)^n \left(\frac{1}{1 - (1 - p)} \right) = (1 - p)^n. \end{aligned}$$

Por lo tanto,

$$p_n = \frac{p(1 - p)^n}{(1 - p)^n} = p \quad \text{para toda } n \geq 0.$$

Podemos partir al contrario, podemos suponer que la tasa de fallos es constante para toda n , i.e $p_n = p \in (0, 1)$ y así llegar a que la única distribución que cumple con esto es la distribución geométrica. Para probar este hecho, procederemos por inducción.

Para $n = 0$ tenemos

$$P(H = 0) = pP(H \geq 0) = p.$$

Ahora para $n = 1$ tenemos

$$P(H = 1) = pP(H \geq 1) = p(1 - P(H = 0)) = p(1 - p).$$

Supongamos que para alguna $n \geq 2$ se satisface que

$$P(H = n) = p(1 - p)^n.$$

Entonces para $n + 1$, tenemos

$$\begin{aligned} P(H = n + 1) &= pP(H \geq n + 1) = p(1 - P(H \leq n)) \\ &= p(1 - P(H = 0) - P(H = 1) - \cdots - P(H = n)) \\ &= p\{1 - p - p(1 - p) - p(1 - p)^2 - \cdots - p(1 - p)^n\} \\ &= p\{1 - p - p[(1 - p) + (1 - p)^2 + \cdots + (1 - p)^n]\} \\ &= p\left\{1 - p - p\left[\frac{1 - p - (1 - p)^{n+1}}{p}\right]\right\} \\ &= p\{1 - p - 1 + p + (1 - p)^{n+1}\} \\ &= p(1 - p)^{n+1}, \end{aligned}$$

con lo que también vemos que se cumple para $n + 1$.

Observemos que el caso clásico del horizonte infinito, $H \equiv \infty$, puede ahora ser considerado como un caso límite de (3.55), haciendo p lo suficientemente pequeño ($p \rightarrow 0$).

Primero que otra cosa necesitamos simplificar ciertas fórmulas para ρ_n^N definidas en (3.35) y (3.36). Es obvio que en este caso de observaciones i.i.d y con $p_k = p$ para toda k

$$\rho_n^N(z) = D\rho_{n+1}^N(z), \quad z \geq 0, \quad (3.56)$$

(ver (3.36)), donde el operador D es definido para cualquier función medible $v : \mathbb{R}^+ \mapsto \mathbb{R}^+$, como

$$Dv(z) = g_\lambda(z)p + (1 - p) \min \{g_\lambda(z), c + Iv(z)\}, \quad (3.57)$$

siendo

$$Iv(z) = E_0 v \left(z \frac{f_1(X)}{f_0(X)} \right), \quad z \geq 0,$$

(ver (3.34)). Debido a que $\rho_N^N(z) = g_\lambda(z)$, $z \geq 0$, tenemos, a partir de (3.56), que

$$\rho_n^N(z) = D^{N-n} g_\lambda(z), \quad z \geq 0, 0 \leq n \leq N. \quad (3.58)$$

Veamos que así es. Aplicando el operador D a la ecuación inmediata anterior tenemos

$$\rho_{N-2}^N(z) = D\rho_{N-1}^N(z) = D(Dg_\lambda(z)) = D^2 g_\lambda(z), \quad z \geq 0.$$

Luego de manera inductiva obtenemos el resultado deseado.

Por lo tanto,

$$\rho_n(z) = \lim_{N \rightarrow \infty} \rho_n^N(z) = \lim_{N \rightarrow \infty} D^N g_\lambda(z), \quad n \geq 0,$$

para toda $z \geq 0$, lo cual significa que

$$\rho_n(z) = \rho(z) = \lim_{N \rightarrow \infty} D^N g_\lambda(z), \quad z \geq 0, \quad (3.59)$$

para cualquiera que sea $n \geq 0$.

Tenemos ahora, a partir de (3.6), que

$$\rho(z) = g_\lambda(z)p + (1-p) \min \{g_\lambda(z), c + I\rho(z)\}, \quad z \geq 0.$$

Debido a que hemos considerado una tasa de fallos constante en cada etapa (horizonte con distribución geométrica), vemos que la anterior función sólo depende de z y ya no de la etapa n .

En este caso los elementos de la regla de paro ahora son $\psi_n \simeq I_{\{g_\lambda(z) \leq c + I\rho(z)\}}$ durante toda etapa n .

Del Lema 3.6 tenemos el siguiente corolario.

Corolario 3.2. *Supongamos que $P(H > 1) > 0$ y la condición de equivalencia (3.54) se cumple. Si*

$$\lambda > c + I\rho(\lambda), \quad (3.60)$$

entonces existen constantes A, B , $0 < A \leq \lambda \leq B$, tales que

$$A = c + I\rho(A) \quad \text{y} \quad \lambda = c + I\rho(B) \quad (3.61)$$

y la desigualdad

$$g_\lambda(z) \leq c + I\rho(z) \quad (3.62)$$

se satisface si y sólo si $z \notin (A, B)$.

El siguiente resultado es una consecuencia directa del Teorema 3.2.

Teorema 3.3. *Supongamos que las condiciones del corolario 3.2 se satisfacen y sean A, B que satisfacen (3.61). Sea (ψ, ϕ) cualquier prueba secuencial tal que*

$$I_{\{z_n \notin [A, B]\}} \leq \psi_n(X_1, \dots, X_n) \leq I_{\{z_n \notin (A, B)\}} \quad P_0\text{-c.s. en } \mathcal{T}_n^\psi$$

para toda $n \geq 0$ y

$$\phi_n(X_1, \dots, X_n) \simeq I_{\{z_n \geq \lambda\}} \quad P_0\text{-c.s. en } \mathcal{S}_n^\psi$$

para toda $n \geq 0$. Entonces para toda prueba secuencial (ψ', ϕ') tal que

$$\alpha(\psi', \phi') \leq \alpha(\psi, \phi) \quad \text{y} \quad \beta(\psi', \phi') \leq \beta(\psi, \phi), \quad (3.63)$$

se cumple que

$$E_0(\tau_\psi \wedge H) \leq E_0(\tau_{\psi'} \wedge H). \quad (3.64)$$

La desigualdad en (3.64) es estricta, si al menos una de las desigualdades en (3.63) es estricta.

Si existen igualdades en todas las desigualdades en (3.63) y (3.64) entonces

$$I_{\{z_n \notin [A, B]\}} \leq \psi'_n(X_1, \dots, X_n) \leq I_{\{z_n \notin (A, B)\}} \quad P_0\text{-c.s. en } \mathcal{T}_n^{\psi'}$$

para toda $n \geq 0$ y

$$\phi'_n(X_1, \dots, X_n) \simeq I_{\{z_n \geq \lambda\}} \quad P_0\text{-c.s. en } \mathcal{S}_n^{\psi'}$$

para toda $n \geq 0$.

Si (3.60) no se cumple, entonces la prueba secuencial con $\psi_0 = 1$ (detener el experimento sin haber tomado observación alguna) y $\phi_0 \simeq I_{\{\lambda \leq 1\}}$ es óptima en el sentido de la primera parte del Teorema 3.3 (entonces la prueba termina sin haber tomado observación alguna). De lo contrario, se sigue del mismo Teorema 3.3 que $\psi_0 = 0$ siempre que $A < \mathcal{Z}_0 < B$, i.e. $A < 1 < B$, que significa que la prueba óptima tiene que iniciar tomando al menos una observación. Si a lo más $A = 1$ o $B = 1$, entonces $\psi_0 \in [0, 1]$ es cualquier número, que es la aleatorización opcional inicial en la etapa 0 y por lo tanto no cuesta nada realizar una prueba así.

Observemos que si la distribución de las hipótesis P_0 y P_1 son diferentes, i.e.

$$\mu(\{x : f_0(x) \neq f_1(x)\}) > 0,$$

entonces integrando bajo H_0 en (3.60), podemos ver que la desigualdad (3.60) se satisface al menos para c lo suficientemente pequeño.

Se sigue del Teorema 3.3 que las pruebas óptimas, ante la presencia de un horizonte aleatorio (3.55), tiene las mismas reglas de paro como la SPRT extendida, introducidas en [Burkholder y Wijsman \(1963\)](#) para el caso clásico (horizonte infinito). Inclusive la regla de decisión coincide con la de la SPRT extendida, cuando el paro ocurre de cierta manera “natural”, que no interviene el horizonte, esto ocurre porque si $\mathcal{Z}_n \geq B$ entonces $\mathcal{Z}_n > \lambda$, lo cual nos lleva a rechazar H_0 , y si $\mathcal{Z}_n \leq A$ entonces $\mathcal{Z}_n < \lambda$, que significa aceptar H_0 . Desafortunadamente, si el paro ocurre al alcanzar primero el horizonte en lugar de τ_ψ ($H < \tau_\psi$), es posible que $\mathcal{Z}_H \in (A, B)$, y la decisión final óptima debe basarse en si $\mathcal{Z}_H$ está por abajo o arriba de $\lambda \in (A, B)$, y no sólo es compararlo con A y/o B . Vemos que la SPRT extendida de [Burkholder y Wijsman \(1963\)](#), además de A y B , adquiere un parámetro adicional, $K = \lambda$, cuando el horizonte aleatorio está presente.

El siguiente teorema nos muestra que si P_0 y P_1 son diferentes, entonces *toda* prueba secuencial con una regla de paro del tipo SPRT extendida tiene la propiedad de optimalidad del Teorema 3.3, con una elección adecuada de cierta constante K . A este tipo de problema le hemos denominado problema inverso, debido a que ahora vamos de manera contraria, ya que fijamos dos valores A, B y ahora debemos determinar un valor constante K (realmente es el valor λ en caso de existir) con el cual la prueba es óptima.

Teorema 3.4 (Problema Inverso). *Sean A y B cualesquiera números positivos, tales que $A \leq 1 \leq B$, $A < B$. Entonces existe K , $A \leq K \leq B$, tal que para toda prueba secuencial (ψ, ϕ) para la cual*

$$I_{\{\mathcal{Z}_n \notin [A, B]\}} \leq \psi_n(X_1, \dots, X_n) \leq I_{\{\mathcal{Z}_n \notin (A, B)\}} \quad P_0\text{-c.s. en } \mathcal{T}_n^\psi \quad (3.65)$$

para toda $n \geq 0$ y

$$\phi_n(X_1, \dots, X_n) \simeq I_{\{\mathcal{Z}_n \geq K\}} \quad P_0\text{-c.s. en } \mathcal{S}_n^\psi$$

para toda $n \geq 0$, tiene la siguiente propiedad de optimalidad.

Para toda prueba secuencial (ψ', ϕ') , tal que

$$\alpha(\psi', \phi') \leq \alpha(\psi, \phi) \quad y \quad \beta(\psi', \phi') \leq \beta(\psi, \phi), \quad (3.66)$$

se cumple

$$E_0(\tau_\psi \wedge H) \leq E_0(\tau_{\psi'} \wedge H). \quad (3.67)$$

La desigualdad en (3.67) es estricta, si al menos una de las desigualdades en (3.66) es estricta.

Si hay igualdades en todas las desigualdades en (3.66) y (3.67) entonces

$$I_{\{Z_n \notin [A, B]\}} \leq \psi'_n(X_1, \dots, X_n) \leq I_{\{Z_n \notin (A, B)\}} \quad P_0\text{-c.s. en } \mathcal{T}_n^{\psi'} \quad (3.68)$$

para toda $n \geq 0$ y

$$\phi'_n(X_1, \dots, X_n) \simeq I_{\{Z_n \geq K\}} \quad P_0\text{-c.s. en } \mathcal{S}_n^{\psi'}$$

para toda $n \geq 0$.

La demostración se deja hasta el Apéndice A.10.

Parece que no hay aparentemente una manera para determinar explícitamente la constante K de la prueba óptima del Teorema 3.4. Es necesario hacer una investigación más profunda para saber si es apropiado usar a $K = 1$, o $K = (A + B)/2$, o $K = \sqrt{AB}$, etc., pueden ser apropiadas para uso práctico. Por otra parte, y debido a que se utiliza una sola vez durante el procedimiento estadístico, (en la etapa final y sólo si el experimento se ve interrumpido por el horizonte aleatorio), su valor exacto no es de esperar que sea de gran importancia durante la realización de la prueba, al menos cuando se espera contar con un número sustancial de observaciones.

Tampoco es claro que si dada cualquier prueba estadística secuencial como en el Teorema 3.4 puede ser óptima en el sentido de Wald y Wolfowitz (1948), i.e. que se pueda minimizar al mismo tiempo $E_0(\tau_{\psi'} \wedge H)$ y $E_1(\tau_{\psi'} \wedge H)$ entre todas las pruebas secuenciales que satisfacen (3.66). Un caso obvio, cuando lo anterior sucede, es precisamente cuando tomamos a $H \equiv \infty$, que en nuestro caso se satisface siempre que $p = 0$ en (3.57), porque K nunca es usada en este caso y la prueba (ψ, ϕ) no es mas que la clásica SPRT (ver Burkholder y Wijsman (1963), Lorden (1980)). Más abajo formulamos precisamente esta propiedad en el Corolario 3.3.

Por otra parte, utilizando $E_1(\tau_{\psi'} \wedge H)$ en lugar de $E_0(\tau_{\psi'} \wedge H)$, como criterio de minimización en el Teorema 3.4, obtenemos la optimalidad de las pruebas (ψ, ϕ) con los mismos A y B , pero posiblemente una diferente K . Si esto nos da la misma K , obviamente, la optimalidad de Wald-Wolfowitz se sostendrá. Por ejemplo, no es difícil ver que en los casos de distribución normal simétrico ($\alpha(\psi, \phi) = \beta(\psi, \phi)$) y binomial considerados en Weiss (1962), K es la misma para ambas pruebas, por lo que ambas poseen la propiedad de optimalidad de Wald-Wolfowitz y en estos casos además se cumple para $p > 0$ ($H < \infty$). De cualquier manera, la pregunta si se cumple la optimalidad Wald-Wolfowitz generalmente en el caso donde $H < \infty$, permanece abierta.

Corolario 3.3. *Supongamos que $H \equiv \infty$ y que se cumple (3.53).*

Sean A y B cualesquiera números positivos, tales que $A \leq 1 \leq B$, $A < B$. Entonces para toda prueba secuencial (ψ, ϕ) que satisface (3.65) para toda $n \geq 0$ y

$$\phi_n(X_1, \dots, X_n) \simeq I_{\{Z_n \geq B\}} \quad P_0\text{-c.s. en } \mathcal{S}_n^\psi$$

para toda $n \geq 0$, tiene la siguiente propiedad de optimalidad.

Para toda prueba secuencial (ψ', ϕ') , tal que

$$\alpha(\psi', \phi') \leq \alpha(\psi, \phi) \quad y \quad \beta(\psi', \phi') \leq \beta(\psi, \phi), \quad (3.69)$$

se cumple

$$E_0\tau_\psi \leq E_0\tau_{\psi'} \quad y \quad E_1\tau_\psi \leq E_1\tau_{\psi'}. \quad (3.70)$$

Ambas desigualdades en (3.70) son estrictas, si al menos una de las desigualdades en (3.69) es estricta.

Si tenemos igualdades en ambas desigualdades en (3.69) y al menos una de las desigualdades en (3.70), entonces ψ'_n satisface (3.68) para toda $n \geq 0$ y

$$\phi'_n(X_1, \dots, X_n) \simeq I_{\{Z_n \geq B\}} \quad P_0\text{-c.s. en } \mathcal{S}_n^{\psi'}$$

para toda $n \geq 0$.

Observemos que si la condición (3.54) se satisface, entonces la frase “ P_0 -casi seguramente” puede ser libremente cambiado por “ P_1 -casi seguramente” en cualquier parte del Corolario 3.3. Además, la afirmación del Corolario 3.3 es de cierta manera más fuerte que la del Teorema 2 en Lorden (1980), debido a que la parte del “sólo si” nos proporciona la estructura de todas las pruebas secuenciales que comparten la misma optimalidad Wald-Wolfowitz (correspondiente a la última parte del Corolario 3.3).

Capítulo 4

Aplicaciones

4.1. Pruebas secuenciales de Bayes para dos hipótesis compuestas

Para esta sección, la idea general en la optimización de las pruebas secuenciales es observar la forma que tiene la función $\mathcal{L}(\psi, \phi)$ dada en (2.3). Una vez que identificamos los elementos que componen la regla de decisión, pasamos a la forma $\mathcal{L}(\psi)$ dada en (2.15), y con esta última expresión identificamos a los elementos que componen la regla de paro ψ_n .

A continuación aplicaremos los resultados del capítulo anterior, principalmente los resultados de la versión del problema Keifer-Weiss modificado, visto en el Capítulo 2, que obtuvimos bajo el horizonte aleatorio a las pruebas bayesianas de dos hipótesis compuestas.

Cabe recordar que en el enfoque frecuentista se trata a los parámetros como fijos pero desconocidos. La diferencia esencial con el enfoque bayesiano es que los parámetros son vistos como variables aleatorias y por ende tienen distribución de probabilidad.

Para plantear este problema, iniciemos con algo de teoría. Supongamos que tenemos una familia paramétrica de funciones de “densidad” conjunta $\{f_\theta(x_1, \dots, x_n), \theta \in \Theta, n \geq 1\}$ donde suponemos que toda $f_\theta(x_1, \dots, x_n)$ es una función medible con respecto a $(\theta, x_1, \dots, x_n)$, para toda $n = 1, 2, \dots$. En esta sección, para la optimización de este tipo de pruebas secuenciales de Bayes, consideraremos el problema de probar $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$, donde Θ_0 y Θ_1 son dos subconjuntos de un espacio de parámetros Θ tales que $\Theta_0 \cap \Theta_1 = \emptyset$.

Vamos a suponer que existe una distribución a priori en el espacio paramétrico, i.e. $\pi \in \Theta$.

Recordemos que bajo toda esta teoría lo novedoso es la presencia del horizonte aleatorio H , el cual es independiente de las observaciones. Para esta sección no impondremos ninguna restricción sobre el tipo de distribución que tenga dicho horizonte.

Definición 4.1. *Dada cualquier $\pi \in \Theta$ y para toda prueba secuencial (ψ, ϕ) , definimos el riesgo bayesiano como*

$$w(\psi, \phi) = c \int_{\Theta} E_{\theta}(\tau_{\psi} \wedge H) d\pi(\theta) + a \int_{\Theta_0} \alpha_{\theta}(\psi, \phi) d\pi(\theta) + b \int_{\Theta_1} \beta_{\theta}(\psi, \phi) d\pi(\theta), \quad (4.1)$$

donde

$$\alpha_\theta(\psi, \phi) = \sum_{n=0}^{\infty} E_\theta \phi_n t_n^\psi h_n^\psi \quad y \quad \beta_\theta(\psi, \phi) = \sum_{n=0}^{\infty} E_\theta (1 - \phi_n) t_n^\psi h_n^\psi$$

para $\theta \in \Theta$, siendo $c > 0$ el costo unitario por observación, $a, b > 0$.

Cabe recordar que el horizonte aleatorio está precisamente inmerso en la función h_n^ψ , donde $h_n^\psi = \psi_n P(H > n) + P(H = n)$ para toda n .

En este contexto estamos pidiendo, sin perder generalidad, que las constantes a y b sean positivas, a diferencia de Novikov (2009) donde pide que las constantes (también conocidas como pesos) que acompañan a las probabilidades de los errores tipo I y tipo II sean no negativas. En caso de no ser así, entonces veremos que la minimización del riesgo bayesiano se vuelve trivial. Supongamos por ejemplo que $a = 0$, entonces en (4.1) estaríamos exentos de cometer el error tipo I, porque no suma en nada a la función riesgo, entonces podemos simplemente tomar siempre la decisión de rechazar H_0 (significa que $\alpha_\theta(\psi, \phi) = 1$), y por consiguiente tendremos que $\beta_\theta(\psi, \phi) = 0$. Por lo tanto, $w(\psi, \phi) = c \int_{\Theta} E_\theta(\tau_\psi \wedge H) d\pi(\theta) = 0$, es decir, la regla de paro óptima es no tomar observación alguna y la regla de decisión, simplemente es $\phi_0 \in [0, 1]$ (la aleatorización inicial). Algo análogo ocurre si $b = 0$.

Ahora es cuando debemos notar que la estructura de la función de riesgo bayesiano (4.1) guarda cierta semejanza con la función de Lagrange dada en la ecuación (2.2), y de hecho más adelante veremos cómo se relacionan para poder aplicar los resultados previos, con el fin de minimizar el riesgo bayesiano. Notemos que b es un peso cualquiera positivo que acompaña a $\beta_\theta(\psi, \phi)$; en cambio el peso que acompaña a $\beta(\psi, \phi)$ en (2.2) lo hemos fijado en 1, que para estos fines también puede ser otro peso cualquiera positivo.

Dada cualquier distribución a priori $\pi \in \Theta$, llamamos a cualquier prueba secuencial bayesiana si su riesgo bayesiano (4.1) es mínimo de entre todas las pruebas secuenciales. En esta sección veremos cómo encontrar este tipo de pruebas secuenciales bayesianas bajo un horizonte aleatorio dada cualquier distribución a priori π .

Usando el Teorema de Fubini, es fácil ver que

$$\begin{aligned} \int_{\Theta_0} \alpha_\theta(\psi, \phi) d\pi(\theta) &= \sum_{n=0}^{\infty} \int_{\Theta_0} \int \phi_n t_n^\psi h_n^\psi f_\theta^n d\mu^n d\pi(\theta) \\ &= \sum_{n=0}^{\infty} \int \phi_n t_n^\psi h_n^\psi \int_{\Theta_0} f_\theta^n d\pi(\theta) d\mu^n \\ &= \pi_0 \sum_{n=0}^{\infty} \int \phi_n t_n^\psi h_n^\psi g_0^n d\mu^n, \end{aligned} \tag{4.2}$$

y

$$\begin{aligned}
\int_{\Theta_1} \beta_\theta(\psi, \phi) d\pi(\theta) &= \sum_{n=0}^{\infty} \int_{\Theta_1} \int (1 - \phi_n) t_n^\psi h_n^\psi f_\theta^n d\mu^n d\pi(\theta) \\
&= \sum_{n=0}^{\infty} \int (1 - \phi_n) t_n^\psi h_n^\psi \int_{\Theta_1} f_\theta^n d\pi(\theta) d\mu^n \\
&= \pi_1 \sum_{n=0}^{\infty} \int (1 - \phi_n) t_n^\psi h_n^\psi g_1^n d\mu^n,
\end{aligned} \tag{4.3}$$

donde $g_i^n \equiv \int_{\Theta_i} f_\theta^n d\pi(\theta) / \pi_i$, $\pi_i = \pi(\Theta_i)$, $i = 0, 1$; pero además,

$$\begin{aligned}
\int_{\Theta} E_\theta(\tau_\psi \wedge H) d\pi(\theta) &= \int_{\Theta_0} \sum_{n=0}^{\infty} \int n t_n^\psi h_n^\psi f_\theta^n d\mu^n d\pi(\theta) + \int_{\Theta_1} \sum_{n=0}^{\infty} \int n t_n^\psi h_n^\psi f_\theta^n d\mu^n d\pi(\theta) \\
&= \pi_0 \sum_{n=0}^{\infty} n \int t_n^\psi h_n^\psi g_0^n d\mu^n + \pi_1 \sum_{n=0}^{\infty} n \int t_n^\psi h_n^\psi g_1^n d\mu^n.
\end{aligned} \tag{4.4}$$

Por lo tanto, ahora el riesgo bayesiano (4.1) es equivalente a

$$w(\psi, \phi) = \sum_{n=1}^{\infty} \int t_n^\psi h_n^\psi (c n g^n + a \phi_n \pi_0 g_0^n + b(1 - \phi_n) \pi_1 g_1^n) d\mu^n, \tag{4.5}$$

donde $g^n = \pi_0 g_0^n + \pi_1 g_1^n = \int_{\Theta} f_\theta^n d\pi(\theta)$. Así, la función $w(\psi, \phi)$ tiene la misma estructura que la función de Lagrange (2.2) (recordemos que por conveniencia tomamos la constante que acompaña a $\beta(\psi, \phi)$ igual a 1, pero puede ser cualquiera), por lo que podemos hacer uso de la herramienta teórica vista anteriormente.

Para poder aplicar los resultados previos a esta nueva función, necesitamos redefinir los elementos clave de la estructura de las pruebas óptimas de la variante del problema Kiefer-Weiss modificado. Para esto último, sea

$$l_N = \min\{\pi_0 a g_0^N, \pi_1 b g_1^N\}, \quad N = 1, 2, \dots \tag{4.6}$$

Definimos así a

$$V_N^N = l_N \tag{4.7}$$

y recursivamente

$$V_n^N = l_n p_n + (1 - p_n) \min\{l_n, c g^n + \int V_{n+1}^N d\mu(x_{n+1})\} \tag{4.8}$$

para toda $n = N - 1, N - 2, \dots, 1$.

Con lo antes definido vemos que más adelante tendremos

$$V_n = \lim_{N \rightarrow \infty} V_n^N \tag{4.9}$$

para toda $n = 1, 2, \dots$; y

$$J_n V_{n+1} = J_n V_{n+1}(x_1, \dots, x_n) = \int V_{n+1}(x_1, \dots, x_{n+1}) d\mu(x_{n+1}), \tag{4.10}$$

para toda $n = 0, 1, 2, \dots$. Entonces, con ayuda del Teorema 2.1, vemos que el Teorema 2.3 se transforma en

Teorema 4.1. *Supongamos que $P(H \geq N) > 0$ para toda $N \geq 1$ y la condición (3.3) se satisface. Entonces*

$$w(\psi, \phi) = \inf_{\psi', \phi'} w(\psi', \phi') \quad (4.11)$$

si y sólo si

$$\phi_n \simeq I_{\{\pi_0 a g_0^n \leq \pi_1 b g_1^n\}} \quad \mu^n\text{-c.t.p. en } S_n^\psi \quad (4.12)$$

para toda $n = 0, 1, 2, \dots$, y ψ satisface

$$\psi_n \simeq I_{\{\min\{\pi_0 a g_0^n, \pi_1 b g_1^n\} \leq c g^n + J_n V_{n+1}\}} \quad \mu^n\text{-c.t.p. en } T_n^\psi \quad (4.13)$$

para toda $n = 0, 1, \dots$.

En caso de que la condición que imponemos sobre el horizonte no se satisfaga, es decir, $P(H \geq N) = 0$, entonces podemos ver el problema en la clase de pruebas truncadas con una N menor

$$N = \max\{n \geq 0 : P(H \geq n) > 0\}.$$

Debido a la misma definición del riesgo bayesiano, tenemos que para cualquier prueba secuencial (ψ, ϕ) que satisfaga las condiciones del Teorema 4.1 es una prueba bayesiana, donde (4.12) y (4.13) nos brinda todas las pruebas secuenciales para dos hipótesis compuestas $H_0 : \theta \in \Theta_0$ y $H_1 : \theta \in \Theta_1$ con $\Theta_0 \cap \Theta_1 = \emptyset$.

4.2. Procesos de Markov a tiempo discreto

En esta parte de la investigación abordaremos el Teorema 2.4 desde otro enfoque, lo aplicaremos a pruebas secuenciales de dos hipótesis simples acerca de la distribución de un proceso de Markov.

Recordando un poco y en términos generales, un proceso de Markov a tiempo discreto, es un proceso con la propiedad que, dado un valor X_n , los valores de X_m , $m > n$, no dependen de los valores de X_k , $k < n$; esto es, la probabilidad de cualquier comportamiento a futuro del proceso, cuando su estado presente se conoce de manera exacta, no es alterada aun con el conocimiento adicional de su comportamiento pasado (Karlin y Taylor, 1975).

Supongamos que, bajo la hipótesis la respectiva H_j , $j = 0, 1$, la “densidad” condicional de X_i , dada X_{i-1} es

$$f_{j,i}(x_i | x_{i-1}) \quad i = 2, 3, \dots,$$

y sea $f_j(x_1)$ la “densidad” inicial, donde todas las densidades son con respecto a alguna medida σ -finita μ . Para este caso tenemos que la función de densidad conjunta de $(X_1, \dots, X_n)$, bajo la hipótesis H_j , es

$$f_j^n = f_j^n(x_1, \dots, x_n) = f_j(x_1) \prod_{i=2}^n f_{j,i}(x_i | x_{i-1}), \quad n = 2, 3, \dots, \quad j = 0, 1. \quad (4.14)$$

Definamos para toda $n = 1, 2, \dots$ la razón de verosimilitud como

$$z_n = \begin{cases} \frac{f_1^n}{f_0^n}, & \text{si } f_0^n(x_1, \dots, x_n) > 0, \\ 0, & \text{en otro caso.} \end{cases}$$

Deseamos minimizar $E_0(\tau_\psi \wedge H)$ bajo las restricciones $\alpha(\psi, \phi) \leq \alpha$ y $\beta(\psi, \phi) \leq \beta$, así que en lo subsecuente construiremos las pruebas secuenciales óptimas para el anterior propósito.

Como hicimos en la anterior sección, necesitamos redefinir las funciones ρ_k^n para toda $k, n = 0, 1, 2, \dots$, bajo este enfoque y que nos ayudarán a alcanzar la optimalidad en este contexto, donde supondremos en todo momento que $z \geq 0$, a menos que se diga lo contrario, y $x \in \mathbb{R}$. Primero hagamos

$$I_k v(z, x) = E_0 v \left(z \frac{f_{1,k+1}(X_{k+1}|x)}{f_{0,k+1}(X_{k+1}|x)}, X_{k+1} \right), \quad z \geq 0. \quad (4.15)$$

Sea

$$\rho_k^0(z, x) = g_\lambda(z) = \min\{\lambda, z\}, \quad (4.16)$$

para toda $k = 0, 1, 2, \dots$; y recursivamente para toda $n = 1, 2, \dots$,

$$\rho_{k-1}^n(z, x) = g_\lambda(z) p_{k-1} + (1 - p_{k-1}) \min\{g_\lambda(z), c + I_{k-1} \rho_k^{n-1}(z, x)\} \quad (4.17)$$

para $k \geq 2$, y

$$\rho_0^n(z, x) = \rho_0^n(z) = g_\lambda(z) p_0 + (1 - p_0) \min\{g_\lambda(z), c + I_0 \rho_1^{n-1}(z, x_1)\}, \quad (4.18)$$

donde

$$I_0 \rho_1^{n-1}(z, x_1) = \int \rho_1^{n-1} \left(z \frac{f_1(x_1)}{f_0(x_1)}, x_1 \right) f_0(x_1) d\mu(x_1)$$

Utilizando el método de inducción hacia atrás, podemos verificar que se cumple

$$V_N^N(x_1, \dots, x_N) = \rho_N^0(z_N, x_N) f_0^N(x_1, \dots, x_N),$$

y

$$V_k^N(x_1, \dots, x_k) = \rho_k^{N-k}(z_k, x_k) f_0^k(x_1, \dots, x_k)$$

para toda $k = N - 1, N - 2, \dots, 1$ (ver (2.23) y (2.24)), y que

$$\begin{aligned} J_{k-1} V_k^N &= \int f_0^{k-1} \rho_k^{N-k} \left(z_{k-1} \frac{f_{1,k}(x_k|x_{k-1})}{f_{0,k}(x_k|x_{k-1})}, x_k \right) f_{0,k}(x_k|x_{k-1}) d\mu(x_k) \\ &= f_0^{k-1} I_{k-1} \rho_k^{N-k}(z_{k-1}, x_{k-1}) \end{aligned} \quad (4.19)$$

(ver Lema 2.3).

Continuando con las mismas ideas de la versión del problema Kiefer-Weiss modificado, podemos ver que se cumple

$$\rho_k^N(z, x) \geq \rho_k^{N+1}(z, x) \geq 0, \quad N = 0, 1, 2, \dots$$

Por lo tanto, existe

$$\lim_{N \rightarrow \infty} \rho_k^N(z, x) =: \rho_k(z, x).$$

Entonces tomando el límite en (4.17), cuando $n \rightarrow \infty$ y aplicando el Teorema de Convergencia Dominada (TCD), tenemos para toda $k \geq 2$:

$$\rho_{k-1}(z, x) = g_\lambda(z)p_{k-1} + (1 - p_{k-1}) \min\{g_\lambda(z), c + I_{k-1}\rho_k(z, x)\} \quad (4.20)$$

y de manera análoga para (4.18) tenemos que

$$\rho_0(z) = g_\lambda(z)p_0 + (1 - p_0) \min\{g_\lambda(z), c + I_0\rho_1(z, x_1)\}. \quad (4.21)$$

De igual forma, tomando el límite en (4.19), cuando $N \rightarrow \infty$, para toda $k \geq 2$ (haciendo uso desde luego del TCD), tenemos

$$J_{k-1}V_k = f_0^{k-1}I_{k-1}\rho_k(z_{k-1}, x_{k-1})$$

y

$$J_0V_1 = I_0\rho_1(1).$$

En el Teorema 2.4 hemos supuesto que la densidad verdadera de las observaciones $X_1, \dots, X_n$ del proceso es f^n , que no necesariamente coincide con alguna de las densidades de las hipótesis; sin embargo, en este caso deseamos minimizar el valor esperado muestral bajo H_0 ($\min E_0(\tau_\psi \wedge H)$), por lo que estamos suponiendo que la distribución verdadera es P_0 , con lo cual debemos sustituir a f^n por f_0^n en dicho teorema. Con lo anterior estamos en condiciones de poder expresar el Teorema 2.4 como:

Teorema 4.2. *Supongamos que se cumplen las condiciones del Teorema 2.3.*

Sea (ψ, ϕ) , cualquier prueba secuencial, tal que la regla ϕ cumple

$$\phi_n \simeq I_{\{\lambda f_0^n \leq f_1^n\}} \quad \mu^n\text{-c.t.p. en } S_n^\psi \quad (4.22)$$

para toda $n = 0, 1, 2, \dots$, y ψ satisface

$$\psi_n \simeq I_{\{g_\lambda(z_n) \leq c + I_n\rho_{n+1}(z_n, x_n)\}} \quad \mu^n\text{-c.t.p. en } T_n^\psi \cap \{f_0^n > 0\} \quad (4.23)$$

para toda $n = 0, 1, \dots$. Entonces para toda prueba secuencial (ψ', ϕ') , tal que

$$\alpha(\psi', \phi') \leq \alpha(\psi, \phi) \quad \text{y} \quad \beta(\psi', \phi') \leq \beta(\psi, \phi) \quad (4.24)$$

se tiene

$$E_0(\tau_\psi \wedge H) \leq E_0(\tau_{\psi'} \wedge H). \quad (4.25)$$

La desigualdad en (4.25) es estricta, si al menos una de las desigualdades en (4.24) es estricta.

Si existen igualdades en todas las desigualdades en (4.24) y (4.25) entonces ϕ' también satisface (4.22) (con ϕ' en lugar de ϕ) para toda $n = 0, 1, 2, \dots$, y ψ' satisface (4.23) μ^n -c.t.p. en $T_n^\psi \cap \{f_0^n > 0\}$ para toda $n = 0, 1, \dots$, (con ψ' en lugar de ψ).

Siguiendo las ideas del Capítulo 1 y fijando a $x \in \mathbb{R}$, podemos ver que las funciones $\rho_k^n(\cdot, x)$ y $\rho_k(\cdot, x)$ son cóncavas, continuas y no decrecientes en $\mathbb{R}^+$ para toda $k = 0, 1, 2, \dots$ y toda $n = 0, 1, 2, \dots$. Debido a esto último, la desigualdad que nos define la regla de paro en la etapa n , $g_\lambda(z_n) \leq c + I_n \rho_{n+1}(z_n, x_n)$, es equivalente a

$$z_n \notin (A_n(x_n), B_n(x_n))$$

con $0 < A_n(x) \leq B_n(x) < \infty$, $n = 1, 2, \dots$

Si el proceso en cuestión es un proceso de Markov homogéneo, i.e.,

$$f_{0,i}(x_i|x_{i-1}) = f_0(x_i|x_{i-1}) \quad \text{y} \quad f_{1,i}(x_i|x_{i-1}) = f_1(x_i|x_{i-1}), \quad (4.26)$$

para toda $i = 2, 3, \dots$; y además la distribución del horizonte aleatorio H es geométrica, con lo cual la tasa de fallos es constante para toda n ($p_n = p$), entonces podemos verificar fácilmente que $\rho_k(z, x) = \rho_1(z, x)$ para toda $k = 2, 3, \dots$. Por lo tanto, $A_n(x) = A(x)$ y $B_n(x) = B(x)$ para toda $n = 1, 2, \dots$

Sólo en el caso de que se cumplan ambas cosas a la vez: la distribución de H que sea geométrica y el proceso de Markov homogéneo; entonces $(A(x), B(x))$ será la forma del intervalo de continuación del experimento. Por lo que si el proceso de Markov no es homogéneo o la distribución del horizonte es distinta de la distribución geométrica, entonces el intervalo de continuación del experimento puede ser distinto en cada etapa n .

4.3. Simulaciones Computacionales

Las aproximaciones numéricas tienen mucha aceptación para comprobar o refutar las hipótesis que se hayan establecido con cierta confianza. Podemos encontrar aproximaciones numéricas en los diseños de ensayos clínicos ([Akong'oLuke, 2008](#)) y otras áreas.

4.3.1. Estimando: $\alpha(\psi, \phi)$, $\beta(\psi, \phi)$, $E_0(H \wedge n)$ y $E_1(H \wedge n)$

En esta sección mostraremos algunos resultados que obtuvimos al realizar una simulación por computadora con la teoría hasta ahora desarrollada y haciendo uso del método Montecarlo. Cabe mencionar que la simulación computacional decidimos hacerla con el software libre [GNU Octave](#).

Principalmente hicimos uso de los resultados del Teorema 3.3 para observaciones i.i.d.. Propusimos analizar un caso discreto y otro continuo: observaciones Bernoulli y exponencial, debido al tiempo máquina. De ahora en adelante y durante esta sección entenderemos que los procedimientos siguientes fueron aplicados para ambos casos: continuo y discreto, a menos que se diga lo contrario.

Supusimos la existencia de un horizonte aleatorio H que con cierta frecuencia interrumpe nuestro experimento y que se rige de manera independiente a las observaciones del proceso, pero con distribución conocida, de hecho geométrica con tasa de fallos $p_n = p$.

Dada una prueba secuencial (ψ, ϕ) , establecimos la prueba de hipótesis simple

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta = \theta_1, \quad \theta_0 < \theta_1.$$

Para iniciar el método secuencial debemos fijar dos números $0 < A < B < \infty$, y que por cuestiones prácticas hemos propuesto que dichas constantes sean las cotas de Wald (1947), esto es $A = \frac{\beta}{1-\alpha}$ y $B = \frac{1-\beta}{\alpha}$. Propusimos el caso simétrico, esto es, cuando $\alpha = \beta = 0.1$, entonces $A = 0.1$ y $B = 9$.

El objetivo general para una prueba como la anterior es estimar las probabilidades de los errores tipo I y II mediante montecarlo y compararlas con el caso simétrico, que denotamos por $\alpha(\psi, \phi)$ y $\beta(\psi, \phi)$, respectivamente. El dominio de las gráficas son los valores propuestos para la tasa de fallos, tomando valores distintos para la tasa de fallos p debida a H . Elegimos dos conjuntos de datos para p , $\mathcal{A} = \{0.0001, 0.0002, \dots, 0.001\}$ y $\mathcal{B} = \{0.01, 0.02, \dots, 0.1\}$ ($\#\mathcal{A} = \#\mathcal{B} = 10$), de tal manera que se viera el cambio cuando H interviene poco (conjunto $\mathcal{A}$) y cuando H interviene frecuentemente (conjunto $\mathcal{B}$) en las $m = 10,000$ simulaciones computacionales de una misma prueba (réplicas de una misma prueba).

También hemos estimado el tamaño promedio muestral bajo H_0 y bajo H_1 , esto es, $E_0[\min(\tau_\psi, H)]$ y $E_1[\min(\tau_\psi, H)]$, respectivamente. Para cada réplica m de una prueba generamos observaciones “aleatorias” x_n en la etapa n , siempre y cuando el experimento no se interrumpa por la regla de paro o bien por el horizonte. Tomamos observaciones i.i.d., con distribución $f_i(x_n) = f(x_n; \theta_i)$ $i = 0, 1$, es decir, las funciones de probabilidad o densidad, según sea el caso, bajo cada hipótesis $i = 0, 1$.

Para cada réplica m o realización de una misma prueba, generamos al inicio un número aleatorio con distribución geométrica de parámetro $p_m \in \mathcal{A}$ (y luego también para $p_m \in \mathcal{B}$), que representa el horizonte aleatorio y el cual puede interrumpir nuestra prueba o realización antes que la regla de paro, registrando dicho evento.

Notemos que algunas pruebas pueden no ser interrumpidas por horizonte sino por la regla de paro $H \geq \tau_\psi$, y entonces aplicamos la decisión final de la teoría clásica secuencial para decidir sobre una de las hipótesis; sin embargo, en caso de que el horizonte llegue antes de la regla de paro, $H < \tau_\psi$, con tasa de fallos p_m , y que por cuestiones prácticas (tiempo computacional), decidimos tomar a $K = 1$ como el criterio K en la regla de decisión en caso de $H < \tau_\psi$ y $z_n \in (A, B)$.

Una observación que debemos hacer en este momento es que las cotas dadas $\alpha = 0.1$ y $\beta = 0.1$ sólo son válidas si no estuviera presente el horizonte aleatorio H , por lo que no se necesitaría estimar los errores. Decidimos tomar estas cotas para, precisamente, observar qué tan cercanos o alejados están nuestros estimados de estas cotas.

Definamos a $z_n = \prod_{i=1}^n \frac{f_1(x_i)}{f_0(x_i)}$ como nuestra razón de probabilidades de la réplica m . Entonces en la observación (etapa) n , la regla es la siguiente:

1. Si $H \geq n$, verificar los siguientes tres puntos.
 - (i) Si $z_n \leq 0.1$, terminamos el experimento y decidimos por H_0 .
 - (ii) Si $z_n \geq 9$, terminamos el experimento y decidimos por H_1 .
 - (iii) Si $0.1 < z_n < 9$, generamos una nueva observación x_{n+1} y regresamos al punto 1.
2. Si $H < n$ terminamos el experimento y

- (i) si $z_n \leq 1$, decidimos por H_0 .
- (ii) Si $z_n > 1$, decidimos por H_1 ,

donde el valor 1 es el valor propuesto para K .

Tanto para el caso discreto como continuo, realizamos dos pruebas de hipótesis; una donde los parámetros hipotetizados $\theta_0 < \theta_1$ estén relativamente cercanos ($|\theta_0 - \theta_1| \leq \min\{\sigma_1, \sigma_0\}$); y otra donde estos parámetros estén relativamente alejados ($|\theta_0 - \theta_1| \geq \min\{\sigma_1, \sigma_0\}$). El motivo de realizar estas dos pruebas fue para confirmar la parte teórica, en cuanto al tiempo máquina, ya que si los parámetros hipotetizados (θ_0, θ_1) se encuentran relativamente cercanos, le tomará más tiempo a nuestro programa “decidir” por alguna hipótesis; mientras que si los parámetros se encuentran relativamente alejados, entonces el tiempo para “decidir” por una hipótesis es más rápido y además para ver qué el efecto que tiene H en ambos casos.

Observaciones con distribución Bernoulli

Para una distribución Bernoulli con parámetro θ , $P(X = 1) = \theta$, sabemos que su varianza está dada por $\sigma^2 = \theta(1 - \theta)$. Las dos pruebas realizadas fueron:

1. $H_0 : \theta = \theta_0 = 0.5$ vs $H_1 : \theta = \theta_1 = 0.53$, $0.03 = \theta_1 - \theta_0 \leq \min\{0.2491, 0.25\}$
2. $H_0 : \theta = 0.5$ vs $H_1 : \theta = 0.8$, $0.3 \geq \min\{0.16, 0.25\}$

Para simplificar el análisis podemos suponer que la réplica en cuestión es $m = 1$, entonces si en la etapa n el valor de z_n se encuentra en la región de continuación, $z_n \in (A, B)$, siempre que el horizonte no haya truncado nuestro experimento, tendremos que

$$\begin{aligned} \frac{\beta}{1 - \alpha} &< z_n = \prod_{i=1}^n \frac{f_1(x_i)}{f_0(x_i)} < \frac{1 - \beta}{\alpha} \\ \frac{\beta}{1 - \alpha} &< \prod_{i=1}^n \frac{\theta_1^{x_i} (1 - \theta_1)^{1-x_i}}{\theta_0^{x_i} (1 - \theta_0)^{1-x_i}} < \frac{1 - \beta}{\alpha} \\ \frac{\beta}{1 - \alpha} &< \left(\frac{\theta_1}{\theta_0}\right)^{\sum_{i=1}^n x_i} \left(\frac{1 - \theta_1}{1 - \theta_0}\right)^{n - \sum_{i=1}^n x_i} < \frac{1 - \beta}{\alpha} \\ \ln\left(\frac{\beta}{1 - \alpha}\right) &< \ln\left(\frac{\theta_1}{\theta_0}\right) \sum_{i=1}^n x_i + \ln\left(\frac{1 - \theta_1}{1 - \theta_0}\right) \left(n - \sum_{i=1}^n x_i\right) < \ln\left(\frac{1 - \beta}{\alpha}\right) \\ \ln\left(\frac{\beta}{1 - \alpha}\right) &< \ln\left(\frac{\theta_1(1 - \theta_0)}{\theta_0(1 - \theta_1)}\right) \sum_{i=1}^n x_i + n \ln\left(\frac{1 - \theta_1}{1 - \theta_0}\right) < \ln\left(\frac{1 - \beta}{\alpha}\right) \end{aligned}$$

De manera que la prueba es básicamente de la forma

$$\ln\left(\frac{\beta}{1 - \alpha}\right) < v_1 \sum_{i=1}^n x_i + nv_2 < \ln\left(\frac{1 - \beta}{\alpha}\right),$$

o bien

$$-2.197 < v_1 \sum_{i=1}^n x_i + nv_2 < 2.197,$$

donde $v_1 = \ln \left(\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)} \right)$ y $v_2 = \ln \left(\frac{1-\theta_1}{1-\theta_0} \right)$ son dos constantes que contienen los parámetros hipotetizados.

Entonces sustituyendo los valores para observaciones de Bernoulli, tenemos que la regla es:

1. Si $H \geq n$, verificar los siguientes tres puntos.
 - (i) Si $v_1 \sum_{i=1}^n x_i + nv_2 \leq -2.197$, terminamos el experimento y decidimos por H_0 .
 - (ii) Si $v_1 \sum_{i=1}^n x_i + nv_2 \geq 2.197$, terminamos el experimento y decidimos por H_1 .
 - (iii) Si $-2.197 < v_1 \sum_{i=1}^n x_i + nv_2 < 2.197$, generamos una nueva observación x_{n+1} y revisamos el punto 1.
2. Si $H < n$ terminamos el experimento y
 - (i) si $v_1 \sum_{i=1}^n x_i + nv_2 \leq 0$, decidimos por H_0 .
 - (ii) Si $v_1 \sum_{i=1}^n x_i + nv_2 > 0$, decidimos por H_1 .

Observaciones con distribución exponencial

Para el caso continuo de una distribución exponencial con parámetro β , la función de densidad es $f(x; \beta) = \frac{1}{\beta} e^{-x/\beta}$ $x > 0$. Sabemos que para una variable $X \sim \text{Exp}(\beta)$, su valor esperado y varianza son: $E(X) = \beta$ y $\text{Var}(X) = \beta^2$, respectivamente. Las pruebas de hipótesis propuestas para este caso continuo fueron:

1. $H_0 : \beta = 5$ vs $H_1 : \beta = 6$, $1 = \beta_1 - \beta_0 \leq \min\{5, 6\}$
2. $H_0 : \beta = 5$ vs $H_1 : \beta = 11$, $6 \geq \min\{5, 11\}$.

La desigualdad de la regla de paro es básicamente la misma que en el caso discreto

$$-2.197 < v_1 \sum_{i=1}^n x_i + nv_2 < 2.197,$$

sólo que $v_1 = \frac{1}{\beta_0} - \frac{1}{\beta_1}$ y $v_2 = \ln \left(\frac{\beta_0}{\beta_1} \right)$. Por lo que el mismo algoritmo computacional sirve para ambos casos.

Hasta ahora hemos establecido cuatro pruebas de hipótesis (dos por cada caso: discreto y continuo), con dos conjuntos de valores para la tasa de fallos p y cuatro valores por estimar en cada prueba

$$(\alpha(\psi, \phi), \beta(\psi, \phi), E_0[\min(\tau_\psi, H)], E_1[\min(\tau_\psi, H)]).$$

Las siguientes gráficas tienen como dominio la tasa de fallos $p \in \mathcal{A}$ o bien $p \in \mathcal{B}$ debida al horizonte aleatorio, que en el caso de que $p \in \mathcal{A}$, significa que el horizonte aleatorio H interviene poco en la interrupción de manera independiente del experimento; mientras que si $p \in \mathcal{B}$, significa que el horizonte interviene más en la interrupción del experimento. Para cada una de las figuras, vemos que hay dos gráficas: las gráficas de la izquierda muestran el tiempo promedio de paro del experimento (ya sea por la regla de paro o por el horizonte H) bajo ambas hipótesis: $E_0[\min\{n, H\}]$ (curva en azul) y $E_1[\min\{n, H\}]$ (curva en negro); mientras que las gráficas de la derecha muestran la estimación de $\alpha(\psi, \phi)$ (curva en azul) y $\beta(\psi, \phi)$ (curva en negro) y contabilizamos las pruebas que no rebazan las respectivas cotas $\alpha = \beta = 0.1$ (recta en rojo), asimismo calculamos el promedio de réplicas que se detienen sólo debido al horizonte aleatorio.

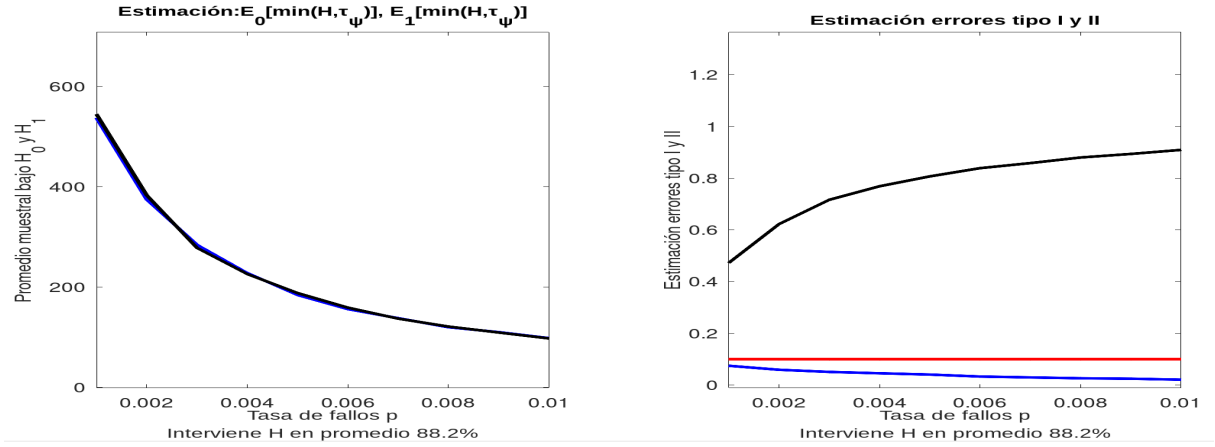


Figura 4.1: Caso Bernoulli, prueba: $H_0 : \theta = 0.5$ vs $H_1 : \theta = 0.53$ y $p \in \mathcal{A}$

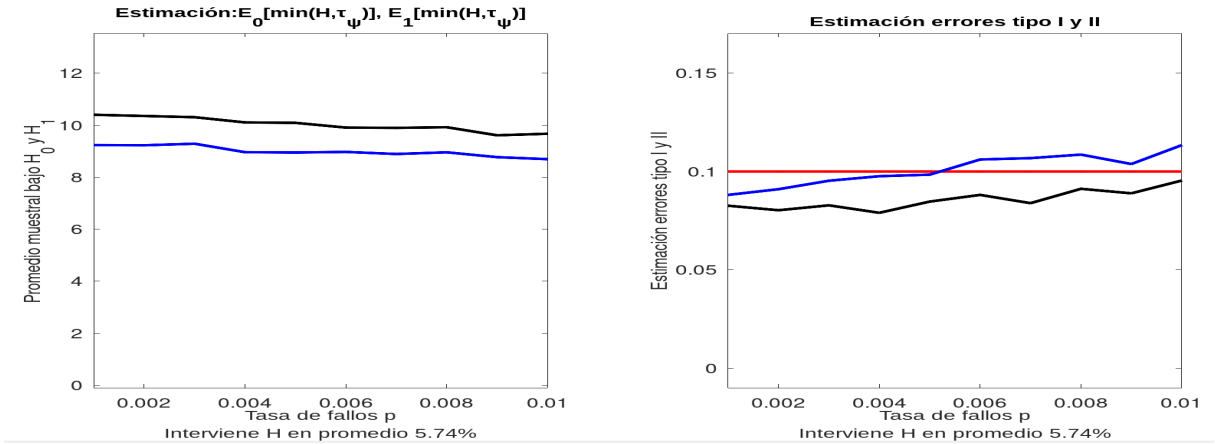


Figura 4.2: Caso Bernoulli, prueba: $H_0 : \theta = 0.5$ vs $H_1 : \theta = 0.8$ y $p \in \mathcal{A}$

Para las primeras 4 figuras, de la 4.1 a la 4.4, el dominio de p es $\mathcal{A}$ y para las 4 figuras

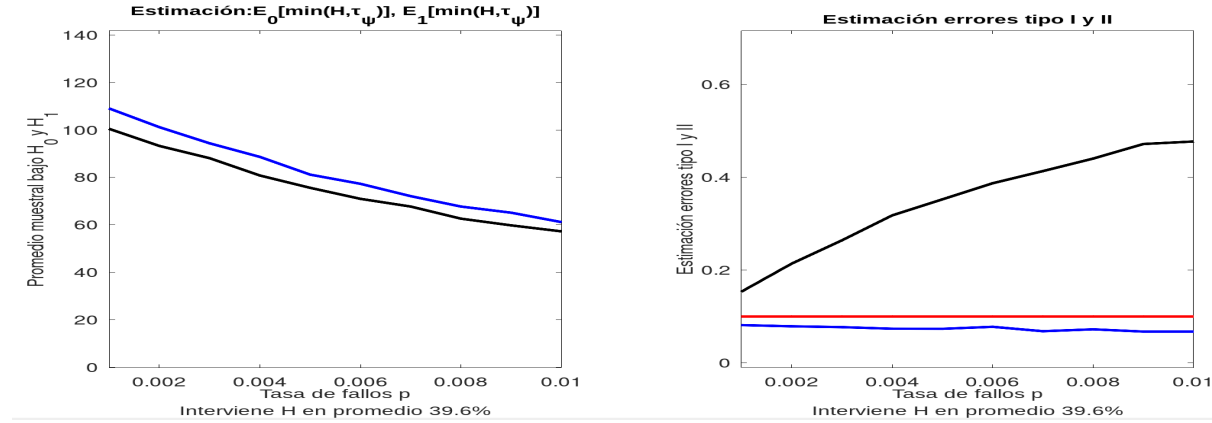


Figura 4.3: Caso exponencial, prueba: $H_0 : \beta = 5$ vs $H_1 : \beta = 6$ y $p \in \mathcal{A}$

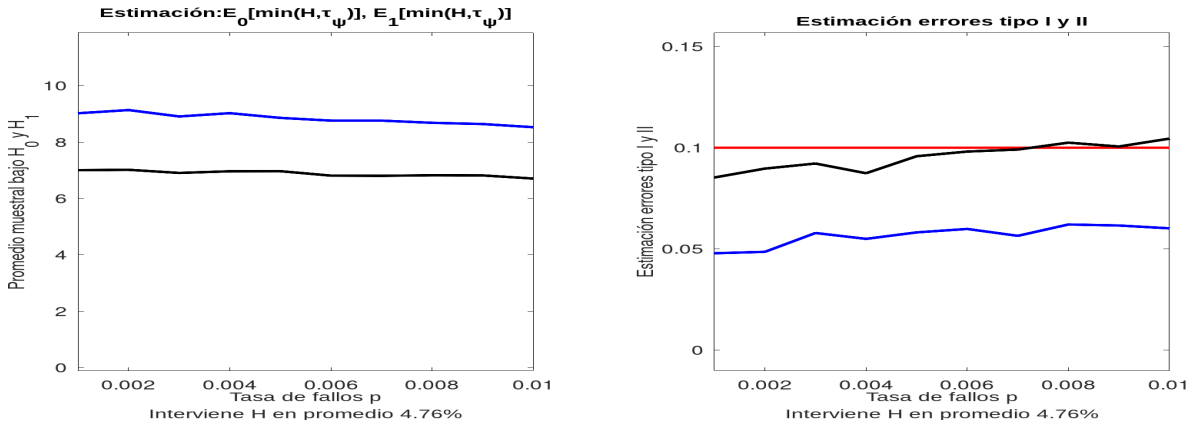


Figura 4.4: Caso exponencial, prueba: $H_0 : \beta = 5$ vs $H_1 : \beta = 11$ y $p \in \mathcal{A}$

restantes, el dominio de p es $\mathcal{B}$. En las figuras 4.1, 4.2, 4.5 y 4.6 las pruebas se hicieron bajo observaciones Bernoulli, mientras que las cuatro restantes se hicieron bajo observaciones exponenciales.

Observemos que en la Figura 4.1 tenemos un tamaño promedio muestral más grande con respecto a las gráficas de las figuras 4.2, 4.3 y 4.4 donde hemos tomado a $p \in \mathcal{A}$. Notemos que si $p \in \mathcal{A}$, entonces $E(H)$ va de 100 a 1000. Por otra parte, la estimación del errores tipo I es el único que está por debajo de la cota y además el horizonte H interviene en alrededor del 88.2% de las veces.

En siguiente Figura 4.2 vemos que el tamaño promedio muestral está alrededor de 10 observaciones, mientras que el error tipo I sigue por debajo de la cota y en algún momento el error tipo II llega a estar por debajo de la cota. El horizonte interrumpe el 5.74% de las veces el experimento.

Para la Figura 4.3 tenemos algo análogo a las primeras dos figuras, pero ahora para observaciones exponenciales. El tamaño promedio muestral bajo ambas hipótesis alrededor

de 100 observaciones. Por otra parte, la estimación de los errores tipo I está por debajo de la cota y el horizonte interrumpe el experimentos en 39.5 % de las veces.

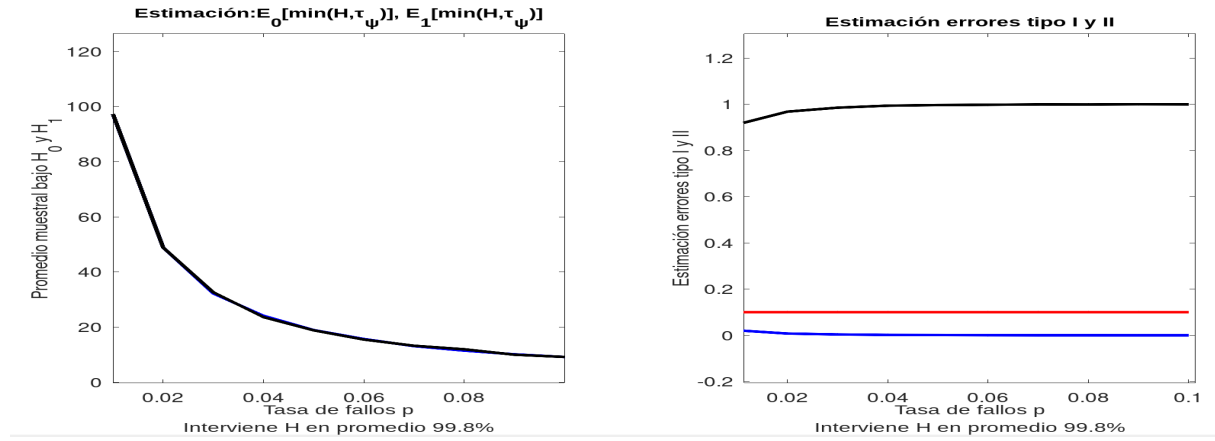


Figura 4.5: Caso Bernoulli, prueba: $H_0 : \theta = 0.5$ vs $H_1 : \theta = 0.53$ y $p \in \mathcal{B}$

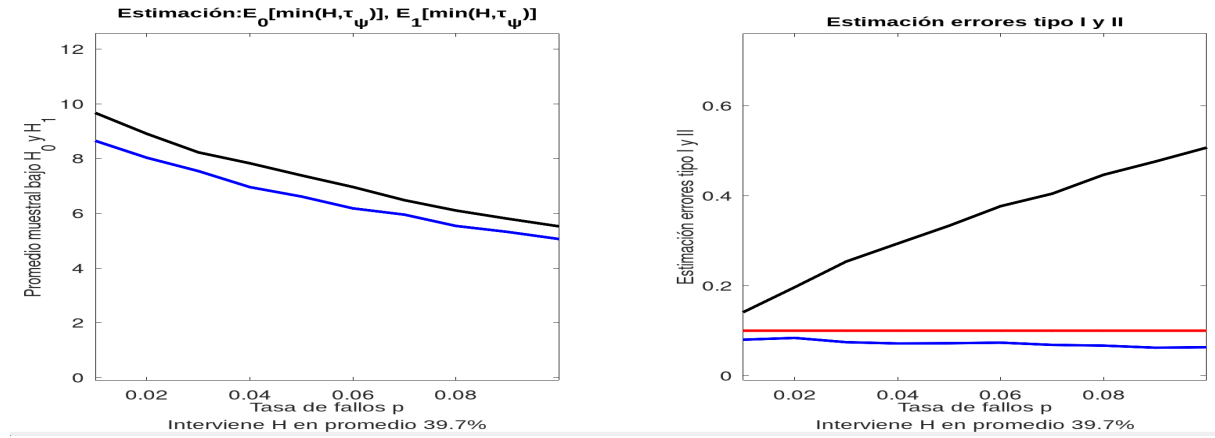


Figura 4.6: Caso Bernoulli, prueba: $H_0 : \theta = 0.5$ vs $H_1 : \theta = 0.8$ y $p \in \mathcal{B}$

En cuanto a la Figura 4.4, el tamaño promedio muestral está alrededor de 8 observaciones y el error tipo I está correcto, donde H interviene en 4.83 % de las veces.

Observe la analogía que tenemos en los resultados de las figuras 4.5-4.8, con respecto a los primeros cuatro resultados, pero ahora $p \in \mathcal{B}$, lo que significa que H interviene más. Tenemos ahora que $E(H)$ va de 10 a 100. Tenemos que el horizonte interviene entre un 35 % y un 90 %. Notemos que los resultados se mejoran un poco si los parámetros hipotetizados se encuentran relativamente alejados.

Es claro que hay más trabajo por hacer en este caso y probar si alguna $K \in (A, B)$ mejora ambos errores cuando el horizonte presenta una tasa de fallos cercana a 0.1. Una propuesta de trabajo futuro es hacer montecarlo con dominio los valores posibles para $K \in$

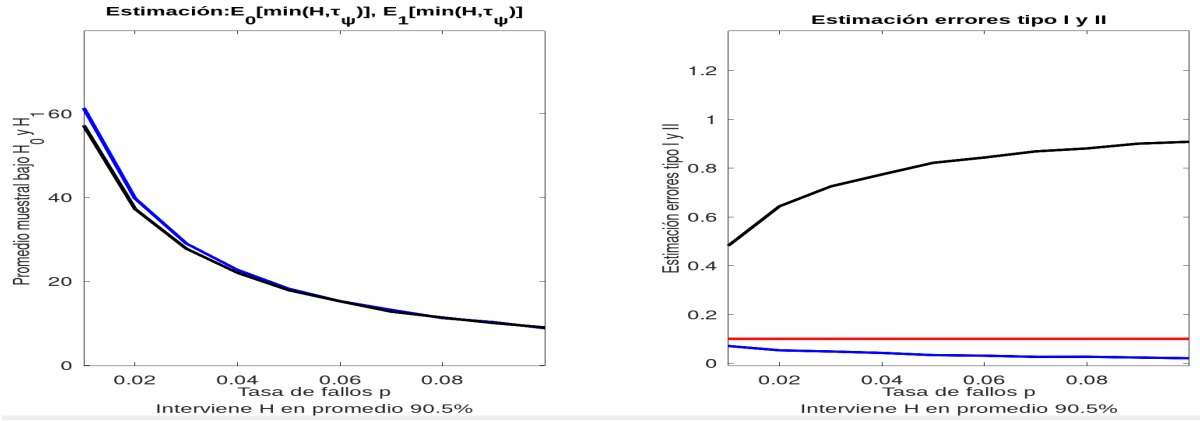


Figura 4.7: Caso exponencial, prueba: $H_0 : \beta = 5$ vs $H_1 : \beta = 6$ y $p \in \mathcal{B}$

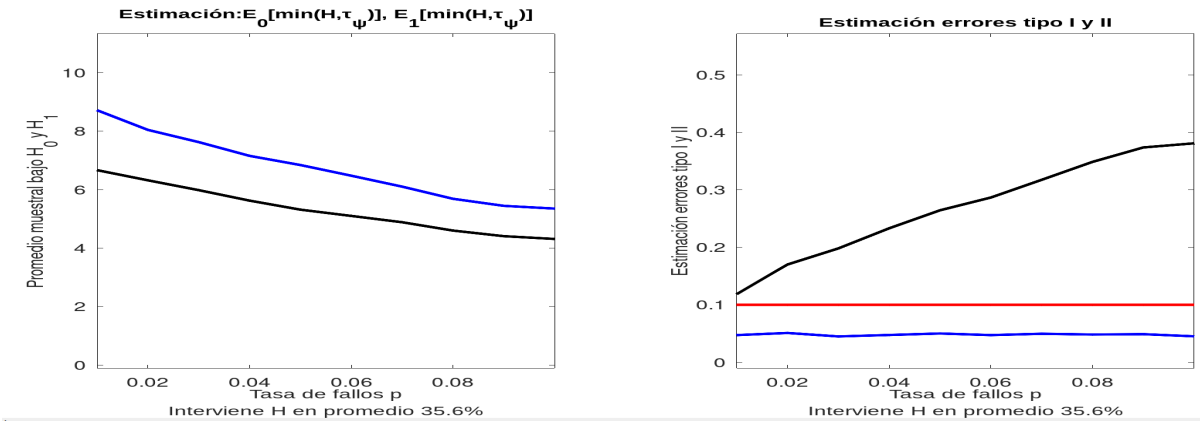


Figura 4.8: Caso exponencial, prueba: $H_0 : \beta = 5$ vs $H_1 : \beta = 11$ y $p \in \mathcal{B}$

(A, B) , porque si $K \rightarrow A$, se mejora la decisión para error tipo II, mientras que si $K \rightarrow B$, la estimación se mejora para el error tipo I. Desde luego que podemos preguntarnos si es posible minimizar el error tipo I y acotar el error tipo II, ¿pero hasta dónde está uno dispuesto a sacrificar el margen de error tipo II?

El código fuente del programa para este análisis lo pueden ver en el Apéndice B.1 y se anexa a la presente investigación con extensión .m, bajo el nombre MonteCarlo.m, el cual puede ser implementado directamente en Matlab si lo desea o bien lo puede descargar directamente de la siguiente [ubicación en onedrive](#).

4.3.2. Obtención de la Región de Continuación

En la sección anterior fijamos el intervalo de continuación ($A = \frac{\beta}{1-\alpha}, B = \frac{1-\beta}{\alpha}$) y supusimos que el horizonte tiene distribución geométrica y estimamos las probabilidades de los errores tipo I y II dadas por la prueba secuencial (ψ, ϕ) , asimismo estimamos

el tamaño muestral promedio bajo ambas hipótesis, tomando la tasa de fallos p en dos conjuntos $\mathcal{A}$ y $\mathcal{B}$. Ahora vamos al problema inverso, esto es, deseamos determinar la región de continuación (A_n, B_n) para cada etapa n , tal como se señala en el Teorema 3.2. Vamos a suponer que estamos en la clase de pruebas truncadas al nivel $N > 0$ (N número máximo de observaciones por tomar y decidir por H_0 o H_1), bajo un horizonte aleatorio H de distribución geométrica. Tomaremos para H una distribución geométrica.

En esta simulación computacional (bajo el método montecarlo) vamos a generar observaciones de Bernoulli i.i.d. para probar

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta = \theta_1, \quad \text{con} \quad \theta_0 < \theta_1, \quad (4.27)$$

donde $\theta_0 = P_0(X = 1)$ y $\theta_1 = P_1(X = 1)$, y fijando $\alpha(\psi, \phi) = \beta(\psi, \phi) = 0.1$ y así determinar la región de continuación (A_n, B_n) para cada etapa n .

Como las observaciones tienen una distribución de Bernoulli y los parámetros hipotetizados son diferentes, entonces se cumplen la condición (3.53).

Vamos a iniciar los cálculos y a guardar algunos resultados de manera inductiva, sólo que la inducción será hacia atrás, esto significa que primero calcularemos y guardaremos los resultados importantes de la etapa N y ya que tengamos los resultados deseados, continuaremos con los cálculos de la etapa $N - 1$ y así hasta llegar a la etapa 1 (programación dinámica).

Para una réplica del experimento secuencial, definamos ahora nuestro estadístico en razón de probabilidades

$$Z_n = Z_n(x_1, x_2, \dots, x_n) = \begin{cases} \frac{f_1^n}{f_0^n}, & \text{si } f_0^n > 0, \\ 0, & \text{en otro caso.} \end{cases}$$

Como f_i^n corresponde a observaciones i.i.d. del tipo Bernoulli, tenemos que

$$f_i^n = \prod_{k=1}^n f_i(x_k) = \prod_{k=1}^n \theta_i^{x_k} (1 - \theta_i)^{1-x_k} = \theta_i^{\sum_{k=1}^n x_k} (1 - \theta_i)^{n - \sum_{k=1}^n x_k}, \quad i = 0, 1$$

Con lo antes mostrado, llegamos a que nuestro estadístico en razón de probabilidades queda como sigue.

$$Z_n = \frac{\theta_1^{\sum_{k=1}^n x_k} (1 - \theta_1)^{n - \sum_{k=1}^n x_k}}{\theta_0^{\sum_{k=1}^n x_k} (1 - \theta_0)^{n - \sum_{k=1}^n x_k}} = \left(\frac{1 - \theta_1}{1 - \theta_0} \right)^n \left(\frac{\theta_1(1 - \theta_0)}{\theta_0(1 - \theta_1)} \right)^{\sum_{k=1}^n x_k}.$$

De esta manera podemos ver que Z_n sólo depende del número de "éxitos" de las observaciones Bernoulli, en otros términos, $Z_n = K_1^n K_2^{\sum_{k=1}^n x_k}$, donde $K_1 = \frac{1 - \theta_1}{1 - \theta_0}$ y $K_2 = \frac{\theta_1(1 - \theta_0)}{\theta_0(1 - \theta_1)}$ son constantes que se conocen a partir de establecer las hipótesis simples del problema. Notemos que en la etapa 1, Z_1 puede tener a lo sumo dos valores distintos $Z_1(X_1) = z(0), z(1)$, donde $z(0) = \frac{1 - \theta_1}{1 - \theta_0}$ o bien $z(1) = \frac{\theta_1}{\theta_0}$; mientras que en la etapa 2 se tiene tres valores distintos, a saber: $Z_2(X_1, X_2) = z(0, 0), z(0, 1) = z(1, 0), z(1, 1)$, donde

$$z(0, 0) = \left(\frac{1 - \theta_1}{1 - \theta_0} \right)^2, \quad z(1, 0) = z(0, 1) = \frac{\theta_1(1 - \theta_1)}{\theta_0(1 - \theta_0)}, \quad z(1, 1) = \left(\frac{\theta_1}{\theta_0} \right)^2.$$

De manera inductiva en la etapa n se tendrá a lo sumo $n + 1$ valores distintos de Z_n correspondientes al número de éxitos $(0, 1, 2, \dots, n)$ y que por supuesto dependerá del número de éxitos que se tenga previamente $(0, 1, 2, \dots, n-1)$ en la etapa anterior $n - 1$. Entonces podemos abusar de la notación y agregar un subíndice a nuestro estadístico con la idea de denotar la suma o número de éxitos en la etapa n -ésima; así $Z_{n,j}$ denotará que en la etapa n hubo j éxitos, donde $j = 0, 1, 2, \dots, n - 1, n$ y $n = 1, 2, \dots, N$. Por lo que ahora $Z_n = Z_{n,j}$, $j = 0, 1, \dots, n$ podemos verlo como un vector de tamaño $n+1$. Sin embargo, para cálculos informáticos vamos a tomar a $Z_{i,j} = Z(i, j) = K_1^i K_2^{j-1}$, donde $i = N, N - 1, \dots, 1$ y $j = 1, 2, \dots, i, i + 1$.

Procediendo como en el Capítulo 4 de observaciones independientes (e idénticamente distribuidas), denotemos como $\rho_{N,j}^N(Z_{N,j}) = g_\lambda(Z_{N,j}) = \min\{\lambda, Z_{N,j}\}$ $j = 0, 1, \dots, N$ el vector de tamaño $N + 1$ que contiene los mínimos entre λ y el valor del estadístico $Z_{N,j}$ (última etapa pero son los primeros cálculos que se guardan) con respecto al j -ésimo éxito. Por lo tanto, en la última etapa N , y para iniciar la inducción hacia atrás, debemos pedir que

$$\begin{aligned}
 \lambda &> c + I_{N-1} \rho_N^N(\lambda) \\
 &= c + \sum_{x=0}^1 \rho_N^N \left(\lambda \frac{P_1(X=x)}{P_0(X=x)} \right) P_0(X=x) \\
 &= c + \sum_{x=0}^1 g_\lambda \left(\lambda \frac{P_1(X=x)}{P_0(X=x)} \right) P_0(X=x) \\
 &= c + \lambda \sum_{x=0}^1 \min\{P_0(X=x), P_1(X=x)\} \\
 &= c + \lambda (\min\{\theta_0, \theta_1\} + \min\{1 - \theta_0, 1 - \theta_1\}) \\
 &= c + \lambda(\theta_0 + 1 - \theta_1).
 \end{aligned}$$

Simplificando el resultado anterior, vemos que la desigualdad anterior es equivalente a

$$\lambda > \frac{c}{\theta_1 - \theta_0}. \quad (4.28)$$

Entonces una vez que introducimos el costo por observación, c , y los parámetros hipotizados, θ_0 y θ_1 , es fácil introducir un valor para $\lambda > 0$ que satisfaga la anterior desigualdad (o equivalentemente, dado λ introducir un valor $c < \lambda(\theta_1 - \theta_0)$). Esto es importante que se cumpla para iniciar la inducción hacia atrás. Desde luego que esta condición sobre el costo por observación tiene justificación porque hemos decidido agregar c al modelo, reemplazando la constante del peso de la probabilidad del error tipo II en la función de Lagrange para continuar teniendo dos grados de libertad y no tres.

Para toda $n = N - 1, N - 2, \dots, 1$ definamos las siguientes funciones:

$$\rho_{n,j}^N(Z_{n,j}) = g_\lambda(Z_{n,j}) p_H + (1 - p_H) \min\{g_\lambda(Z_{n,j}), c + I_n \rho_{n+1}^N(Z_{n,j})\}, \quad (4.29)$$

vectores de tamaño $n + 1$ ($j = 0, 1, \dots, n$), donde para cada $n = N - 1, \dots, 1$ y j fijos,

$$\begin{aligned} I_n \rho_{n+1,j}^N(Z_{n,j}) &= \sum_{x=0}^1 \rho_{n+1,j}^N \left(Z_{n,j} \frac{P_1(X=x)}{P_0(X=x)} \right) P_0(X=x) \\ &= \rho_{n+1,j}^N \left(Z_{n,j} \frac{1-\theta_1}{1-\theta_0} \right) (1-\theta_0) + \rho_{n+1,j+1}^N \left(Z_{n,j} \frac{\theta_1}{\theta_0} \right) \theta_0. \\ &= \rho_{n+1,j}^N(Z_{n+1,j})(1-\theta_0) + \rho_{n+1,j+1}^N(Z_{n+1,j+1})\theta_0, \end{aligned}$$

y el vector $\rho_{n+1,j}^N(Z_{n+1,j})$ $j = 0, 1, \dots, n+1$ ya ha sido previamente calculado y guardado, por lo que los elementos de la etapa n quedan bien identificados para invocar a los resultados de la etapa $n + 1$.

Lo que sigue es encontrar los intervalos de continuación (A_n, B_n) para cada una de las etapas $n = N, N - 1, \dots, 1$. Para dicho propósito es necesario utilizar el resultado del Lema 3.6 y ver qué valores $Z_{n,j}$ cumplen con la condición

$$g_\lambda(Z_{n,j}) \leq c + I_n \rho_{n+1,j}^N(Z_{n,j}) \quad n = N - 1, \dots, 1, j = 1, 2, \dots, n + 1. \quad (4.30)$$

Lo anterior significa que si

$$g_\lambda(Z_{n,j}) > c + I_n \rho_{n+1,j}^N(Z_{n,j}) \quad n = N - 1, \dots, 1, j = 1, 2, \dots, n + 1, \quad (4.31)$$

entonces $Z_{n,j}$ se encontrará en la región de continuación para $n = N - 1, \dots, 1$ y $j = 1, \dots, n + 1$. De esta manera tenemos que el intervalo de continuación es

$$(A_n, B_n) = \left(\min_j \{Z_{n,j}\}, \max_j \{Z_{n,j}\} \right)$$

El vector $Z_{n,j}$, para cada $n = 1, 2, \dots, N - 1$, contiene todas las posibilidades de nuestra razón de probabilidades en la etapa n y que hemos llamado *frontera de observaciones*. La región de continuación (A_n, B_n) está ligada con el mínimo y máximo del vector $Z_{n,j}$, respectivamente, sobre el número de éxitos j en la etapa n , por lo que ya es posible dibujar nuestra región de continuación dentro de la frontera de observaciones.

Una vez que tenemos nuestra región de continuación, podemos generar 10000 réplicas de una muestra de tamaño N de observaciones Bernoulli bajo H_i , $i = 0, 1$. Observemos que $Z_n = K_1^n K_2^{\sum_{k=1}^n x_k} = Z_{n-1} K_1 K_2^{x_n}$, donde $x_n \sim \text{Ber}(\theta_i)$, por lo tanto podemos ver el camino de cada réplica y registrar en qué etapa n se sale de nuestra región de continuación para contabilizar las decisiones correctas al saber que las observaciones que generamos fueron bajo H_i y si el camino fue en esa dirección.

Para cada una de las réplicas inicializamos la variable $s = 1$ y para cada etapa $n = 1, \dots, N$, tenemos la siguiente regla:

- (i) Hacemos $s = s K_1 K_2^{x_n}$.
- (ii) Si $s \leq A_n$, decidimos por H_0 .
- (iii) Si $s \geq B_n$, decidimos por H_1 .

(iv) Si $A_n < s < B_n$, repetir paso (i) con $n + 1$.

De esta manera tenemos identificados el porcentaje de decisiones correcta y podemos también graficar el camino de cada réplica.

Debido a la naturaleza de la razón de probabilidades, es conveniente tomar logaritmos, toda vez que $Z_{n,j}$ puede tomar valores extremadamente pequeños (inclusive por abajo del épsilon de la máquina) o bien extremadamente grandes (no se puede interpretar la gráfica de manera adecuada). Entonces los resultados del programa que determina (y dibuja) la región de continuación de un experimento con observaciones de Bernoulli para probar (4.27) podemos verlos en de las figuras 4.10 a la 4.24.

Para entender esta simulación computacional, revisemos la Figura 4.9, en ella mostramos sólo 10 réplicas del experimento y fijamos el máximo de observaciones por tomar en cada réplica de $N = 32$ (abscisas), pero en las demás gráficas este máximo muestral lo cambiamos a placer. Sólo hemos analizado observaciones Bernoulli y hemos establecido una prueba de hipótesis simple: $H_0 : \theta = 0.5$ s $H_1 : \theta = 0.53$ (en otras ocasiones $\theta_1 = 0.8$). Las líneas en rojo son la frontera de observaciones, esto significa que la razón de probabilidades Z_n (ordenadas) para cualquier etapa n no puede dar un valor por arriba o por abajo de dicha línea. Sin embargo, la curva en azul acota la región de continuación, es decir que si Z_n es nuestra razón de probabilidad y cae dentro de esta región en la etapa n , debemos tomar una nueva observación y continuar con el algoritmo secuencial. Las líneas en verde son los caminos de Z_n en cada réplica. Una vez que Z_n toca la frontera A_n o B_n en la etapa n (línea en azul) se termina la réplica y se decide por H_0 o bien por H_1 de acuerdo a la regla de decisión. Las observaciones de Bernoulli las generamos bajo ambas hipótesis y vemos el porcentaje de correctas decisiones en cada corrida del experimento.

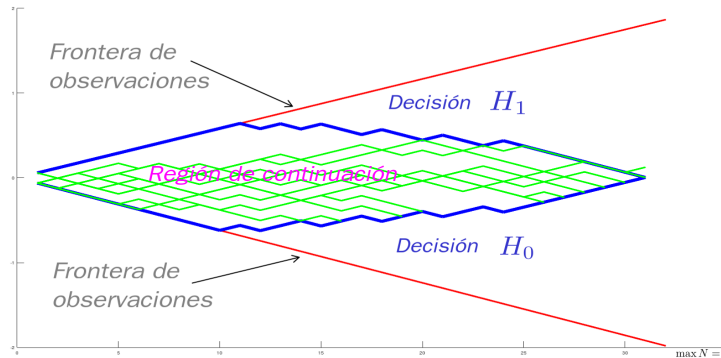


Figura 4.9: Observaciones de Bernoulli (10 réplicas) con parámetro $\theta_0 = 0.5$, $p_H = 0.0001$, $N = 32$, de las cuales 9 decisiones fueron correctas y $E_0(H \wedge n) = 26.6$.

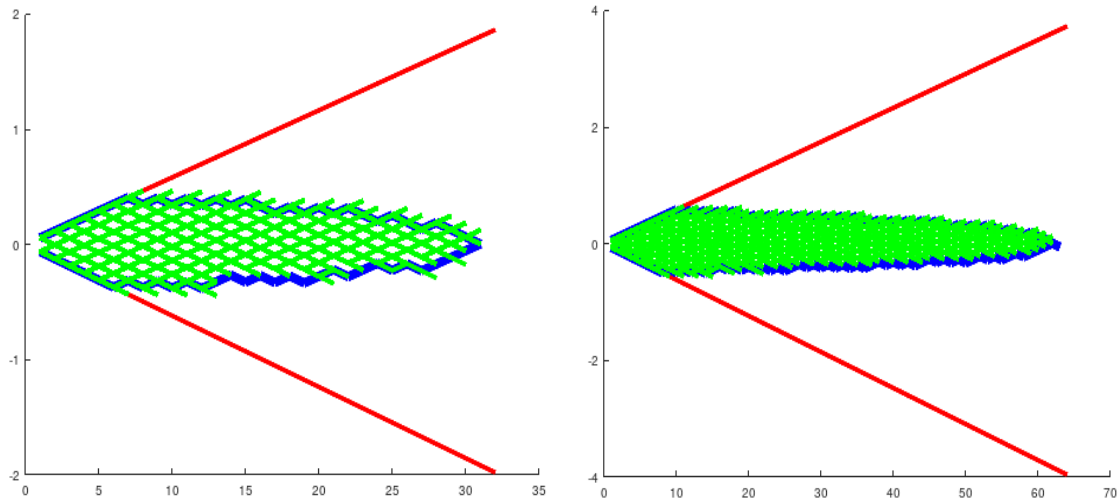
El código fuente del programa que dibuja la región de continuación bajo horizonte, lo pueden ver en el Apéndice B.2 El nombre que hemos asignado a la función en Octave para estos propósitos es **RegionDeContinuacion** (o [link de descarga en onedrive](#)), la cual se invoca como `RegionDeContinuación(N, p_n, \theta_0, \theta_1, H_i)`, donde N es el máximo de observaciones a tomar, p_n es la tasa de fallos, θ_i es el parámetro hipotetizado bajo la hipótesis i y con $H_i = 0, 1$ estaremos indicando bajo qué hipótesis vamos a generar las

observaciones para los caminos. Ya hemos fijado $c = 0.0001$ y $\lambda = 1$, para que de esta manera se puedan realizar los cálculos numéricos.

En el siguiente conjunto de resultados, veremos algunas imágenes en donde la prueba de hipótesis establecida es $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$, $c = 0.001$ y $\lambda = 1$, donde $\theta_0 = 0.5$ y $\theta_1 = 0.53$ o bien $\theta_1 = 0.8$. En teoría el valor del costo c puede ser arbitrario, por lo que no debería estar sujeto a ninguna condición, pero recordemos que c debe cumplir la condición del corolario 3.2 para poder hacer la inducción hacia atrás. Es posible, desde luego, que no se introduzca el valor c en la prueba y al final de una prueba secuencial sólo multiplicar el tamaño promedio muestral por dicho costo observacional para así obtener el costo del experimento. También tenemos otras variantes: $p_H = 0.001$ o bien $p = 0.00001$; además el tamaño máximo para cada réplica fue con $N = 32, 64, 128, 500$ y 1024 . Al final señalamos el porcentaje de buenas decisiones para cada resultado.

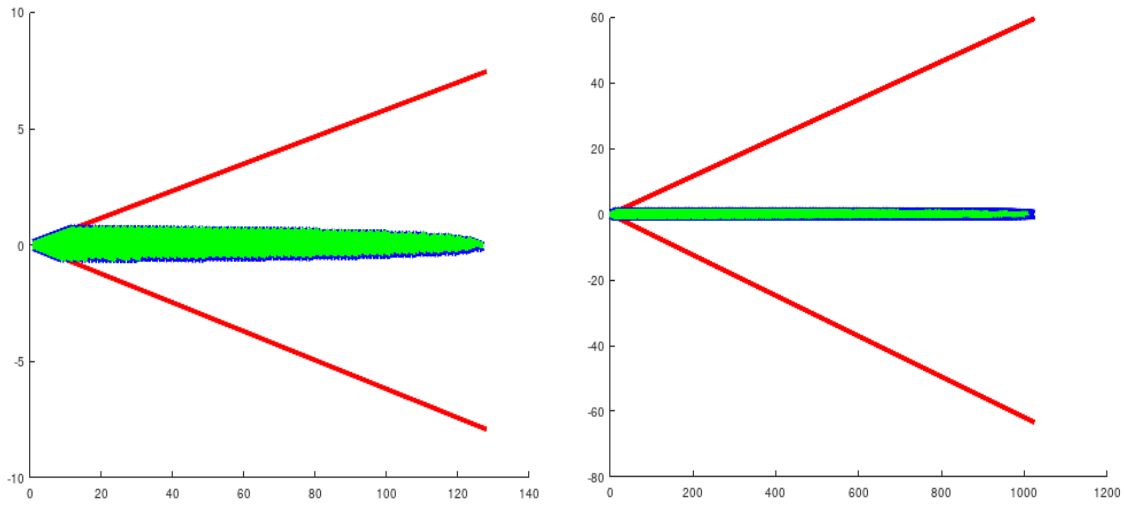
Iniciemos con la Figura 4.10 a), en ella vemos que $N = 32$, $E_0(\min\{H, n\}) = 21.29$, lo que nos indica que las observaciones generadas para esta gráfica fueron bajo la hipótesis nula, obteniendo un 64.6 % de decisiones correctas. En el inciso b), podemos ver que sólo cambiamos $N = 64$, con lo que el tamaño promedio muestral bajo H_0 aumentó 38.35 y las decisiones correctas fueron del 66.12 % y usando $p_H = 0.001$ en ambos casos. Y en la Figura 4.11 las condiciones son iguales, sólo que el tamaño máximo posible es de 128 para a) y 1024 para b). Notemos que en estas últimas gráficas las decisiones correctas aumentaron. Para el siguiente bloque de figuras 4.12 y 4.13 lo que cambia es la tasa de fallos a $p_H = 0.00001$, sin embargo las decisiones correctas se encuentran dentro del 60 % y 80 %.

En la figuras 4.14 y 4.15 es análogo a las figuras



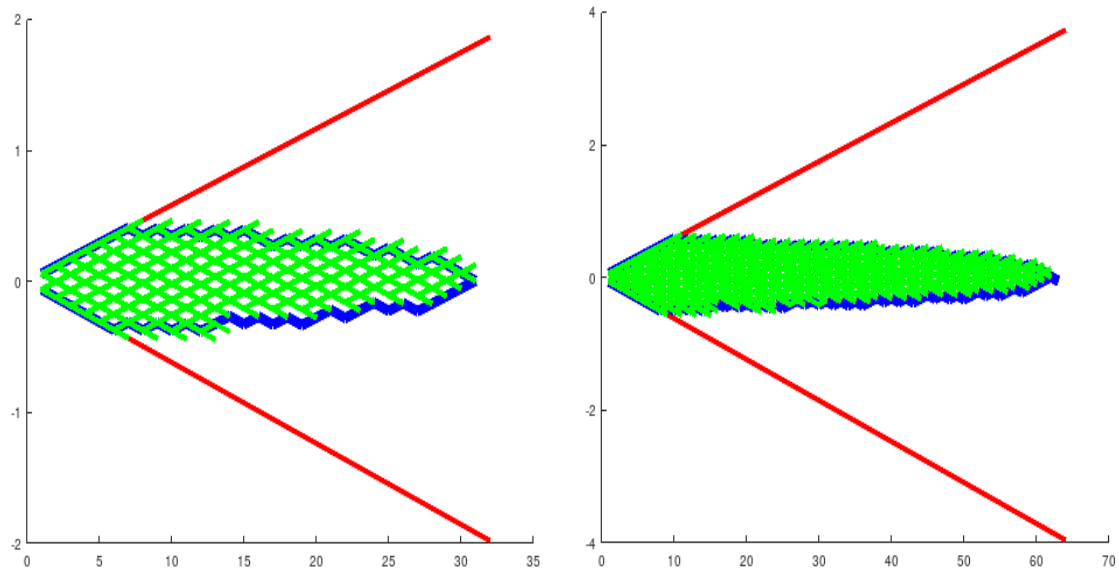
(a) $N = 32$, $E_0(\min\{H, n\}) = 21.29$ y 64.6 % (b) $N = 64$, $E_0(\min\{H, n\}) = 38.35$ y 66.12 % decisiones correctas.

Figura 4.10: Observaciones de Bernoulli bajo $\theta_0 = 0.5$ (vs $\theta_1 = 0.53$) y tasa de fallos $p_H = 0.001$.



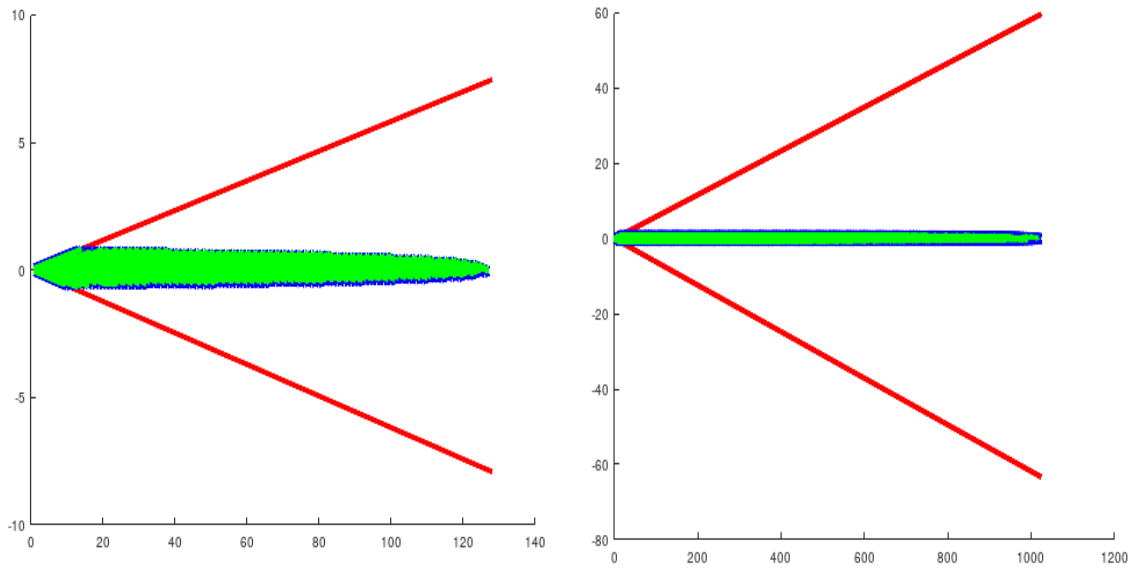
(a) $N = 128$, $E_0(\min\{H, n\}) = 66.74$ y (b) $N = 1024$, $E_0(\min\{H, n\}) = 147.75$ y
 69.34 % decisiones correctas. 77.9 % decisiones correctas.

Figura 4.11: Observaciones de Bernoulli bajo $\theta_0 = 0.5$ (vs $\theta_1 = 0.53$) y tasa de fallos $p_H = 0.001$.



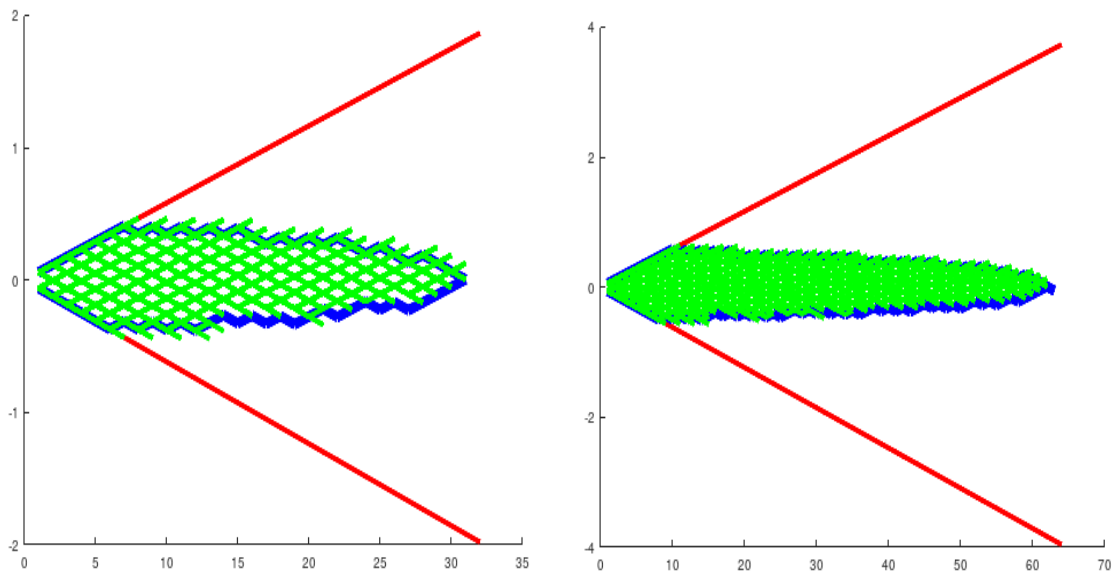
(a) $N = 32$, $E_0(\min\{H, n\}) = 21.11$ y (b) $N = 64$, $E_0(\min\{H, n\}) = 38.47$ y
 65.35 % decisiones correctas. 66.73 % decisiones correctas.

Figura 4.12: Observaciones de Bernoulli bajo $\theta_0 = 0.5$ (vs $\theta_1 = 0.53$) y tasa de fallos $p_H = 0.00001$.



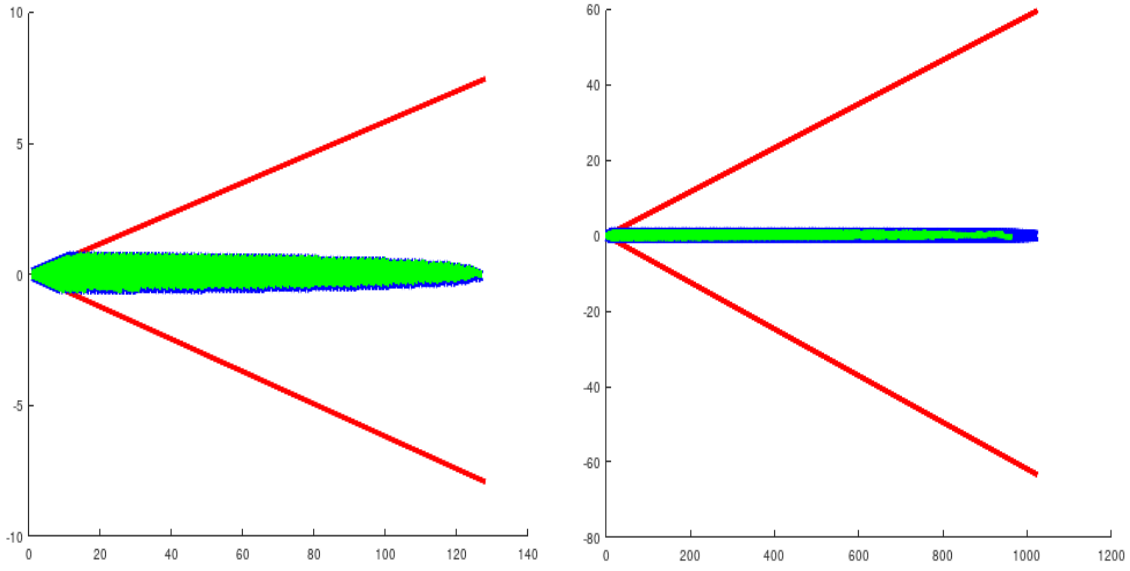
(a) $N = 128$, $E_0(\min\{H, n\}) = 67.94$ y 69.28% decisiones correctas. (b) $N = 1024$, $E_0(\min\{H, n\}) = 166.1$ y 78.94% decisiones correctas.

Figura 4.13: Observaciones de Bernoulli bajo $\theta_0 = 0.5$ (vs $\theta_1 = 0.53$) y tasa de fallos $p_H = 0.00001$.



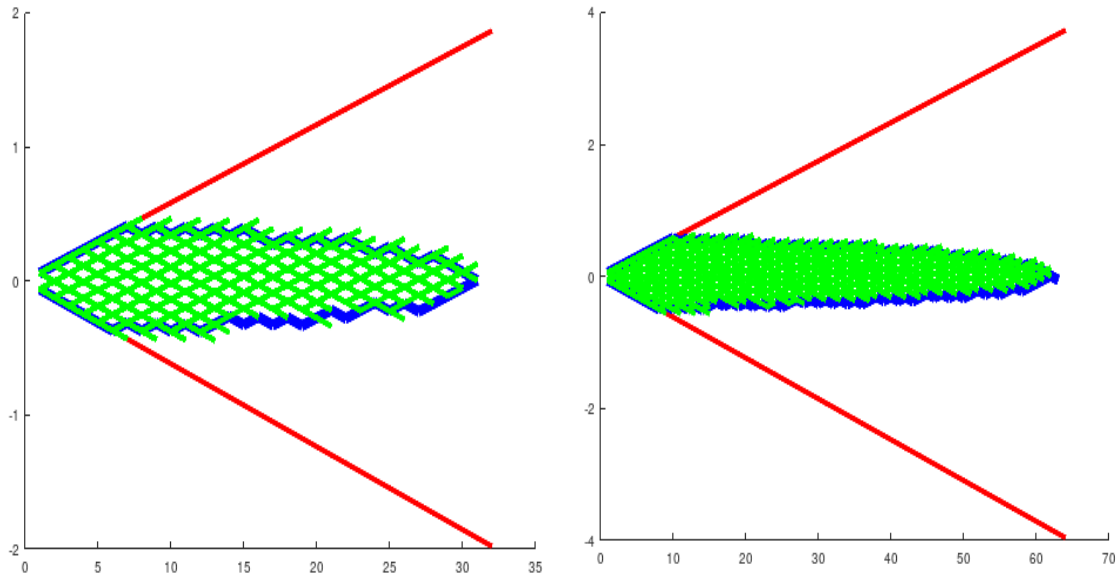
(a) $N = 32$, $E_1(\min\{H, n\}) = 21.73$ y 47.45% decisiones correctas. (b) $N = 64$, $E_1(\min\{H, n\}) = 40$ y 51.4% decisiones correctas.

Figura 4.14: Observaciones de Bernoulli bajo $\theta_1 = 0.53$ (vs $\theta_1 = 0.5$) y tasa de fallos $p_H = 0.001$.



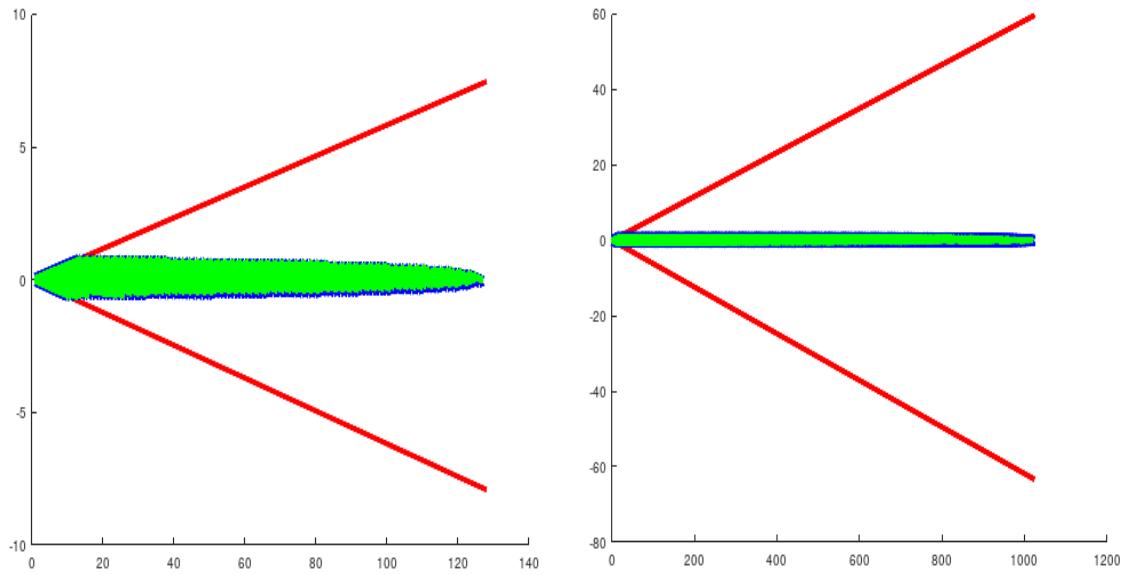
(a) $N = 128$, $E_1(\min\{H, n\}) = 71.23$ y (b) $N = 1024$, $E_1(\min\{H, n\}) = 164.72$ y
 54.54 % decisiones correctas. 57.52 % decisiones correctas.

Figura 4.15: Observaciones de Bernoulli bajo $\theta_1 = 0.53$ (vs $\theta_1 = 0.5$) y tasa de fallos $p_H = 0.001$.



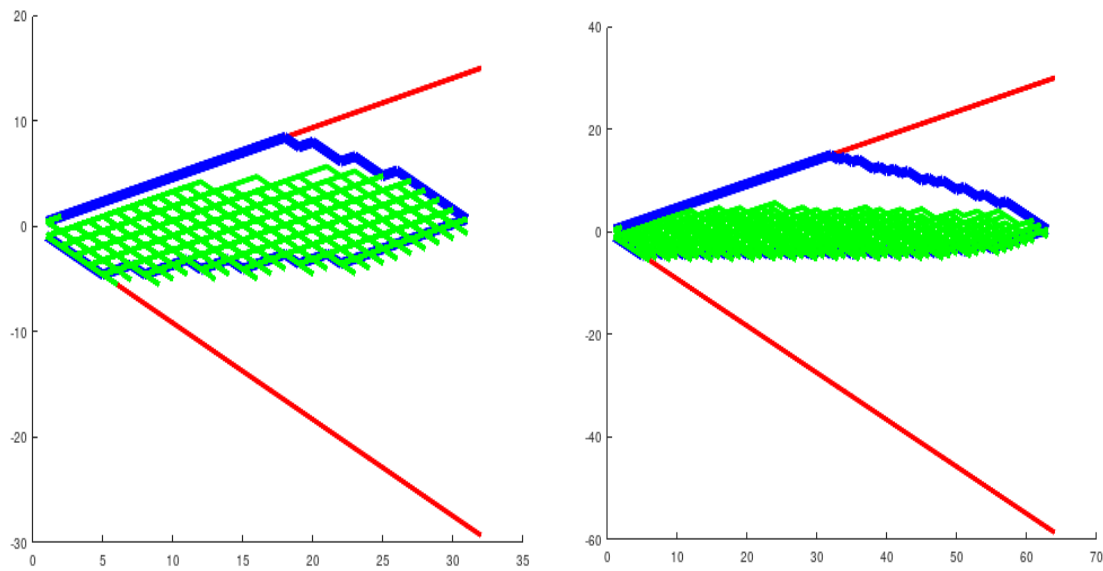
(a) $N = 32$, $E_1(\min\{H, n\}) = 21.28$ y (b) $N = 64$, $E_1(\min\{H, n\}) = 40.13$ y
 47.63 % decisiones correctas. 51.55 % decisiones correctas.

Figura 4.16: Observaciones de Bernoulli bajo $\theta_1 = 0.53$ (vs $\theta_1 = 0.5$) y tasa de fallos $p_H = 0.00001$.



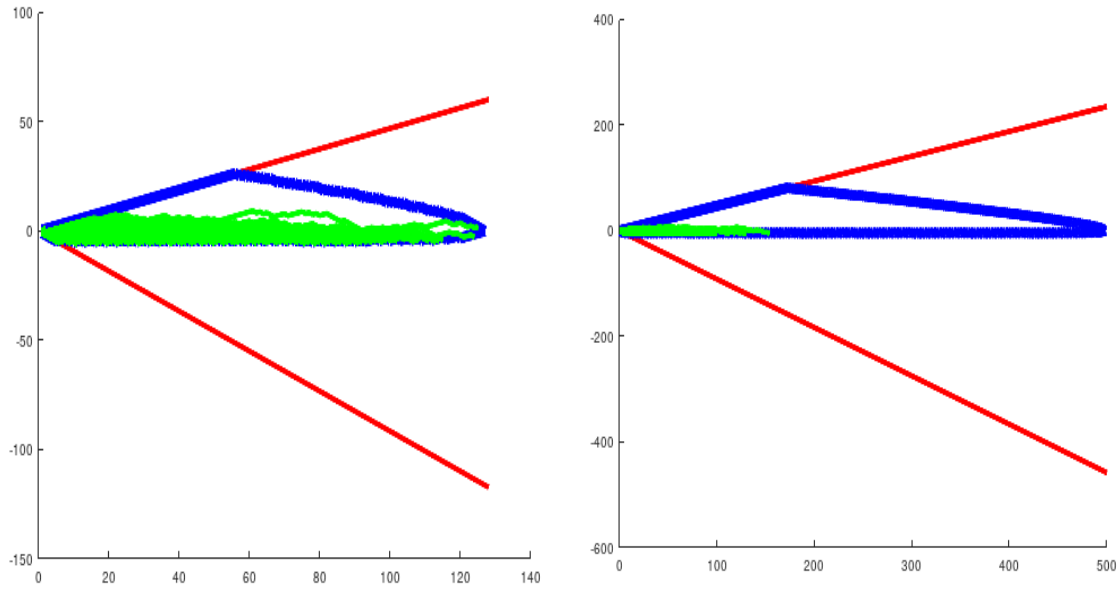
(a) $N = 128$, $E_1(\min\{H, n\}) = 71.88$ y 54.79% decisiones correctas. (b) $N = 1024$, $E_1(\min\{H, n\}) = 186.28$ y 60% decisiones correctas.

Figura 4.17: Observaciones de Bernoulli bajo $\theta_1 = 0.53$ (vs $\theta_1 = 0.5$) y tasa de fallos $p_H = 0.00001$.



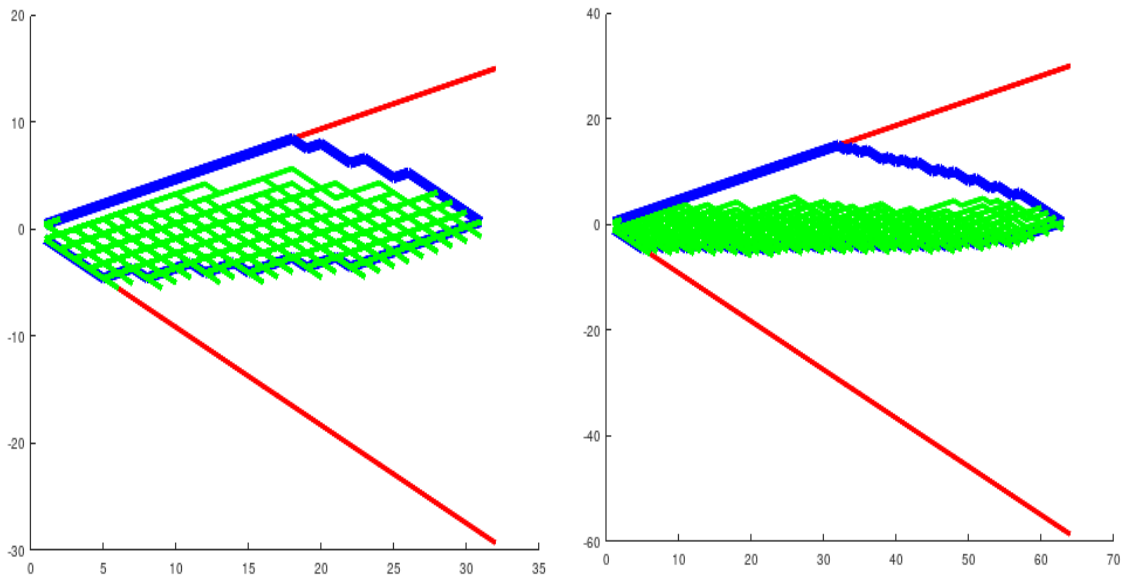
(a) $N = 32$, $E_0(\min\{H, n\}) = 12.66$ y 72.8% decisiones correctas. (b) $N = 64$, $E_0(\min\{H, n\}) = 15.46$ y 74.17% decisiones correctas.

Figura 4.18: Observaciones de Bernoulli bajo $\theta_0 = 0.5$ (vs $\theta_1 = 0.8$) y tasa de fallos $p_H = 0.001$.



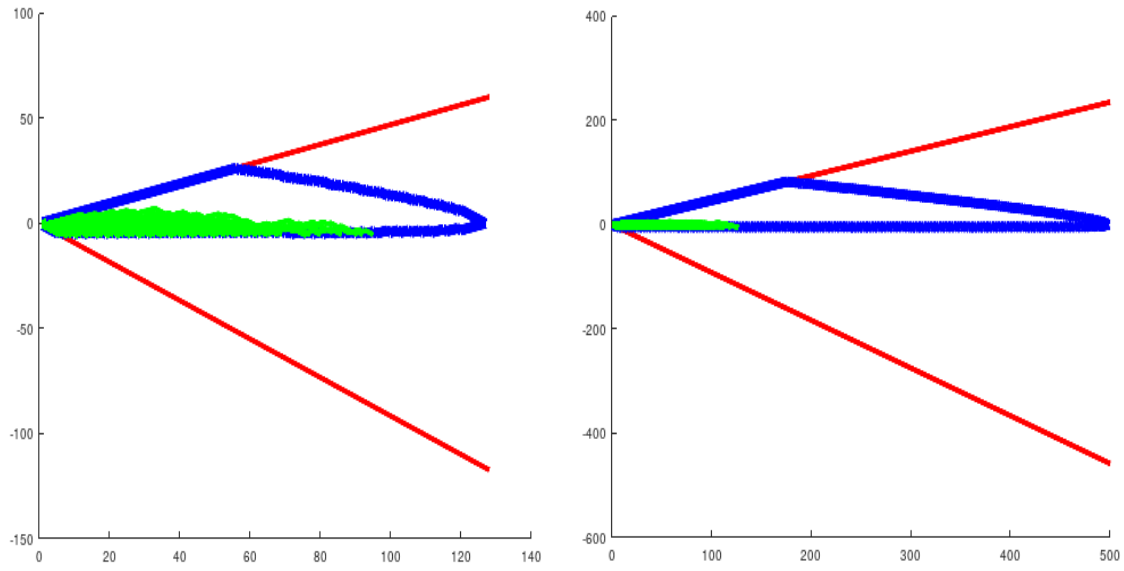
(a) $N = 128$, $E_0(\min\{H, n\}) = 17.48$ y (b) $N = 500$, $E_0(\min\{H, n\}) = 17.43$ y
75.55 % decisiones correctas. 75.3 % decisiones correctas.

Figura 4.19: Observaciones de Bernoulli bajo $\theta_0 = 0.5$ (vs $\theta_1 = 0.8$) y tasa de fallos $p_H = 0.001$.



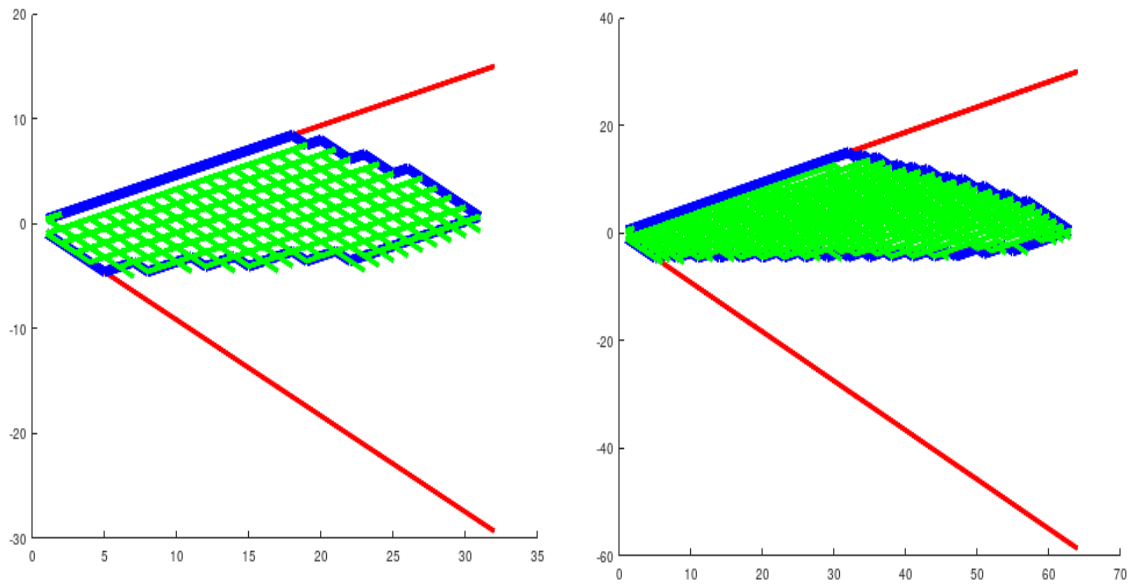
(a) $N = 32$, $E_0(\min\{H, n\}) = 12.69$ y (b) $N = 64$, $E_0(\min\{H, n\}) = 15.87$ y
72.85 % decisiones correctas. 75.02 % decisiones correctas.

Figura 4.20: Observaciones de Bernoulli bajo $\theta_0 = 0.5$ (vs $\theta_1 = 0.8$) y tasa de fallos $p_H = 0.00001$.



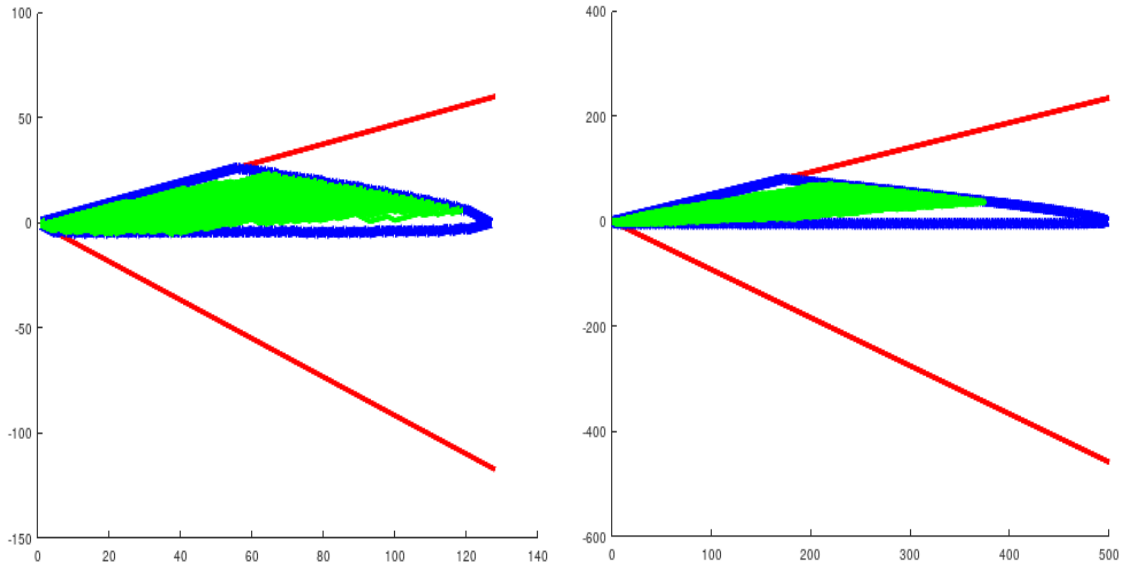
(a) $N = 128$, $E_0(\min\{H, n\}) = 17.37$ y (b) $N = 500$, $E_0(\min\{H, n\}) = 17.58$ y
 72.62 % decisiones correctas. 74.8 % decisiones correctas.

Figura 4.21: Observaciones de Bernoulli bajo $\theta_0 = 0.5$ (vs $\theta_1 = 0.8$) y tasa de fallos $p_H = 0.00001$.



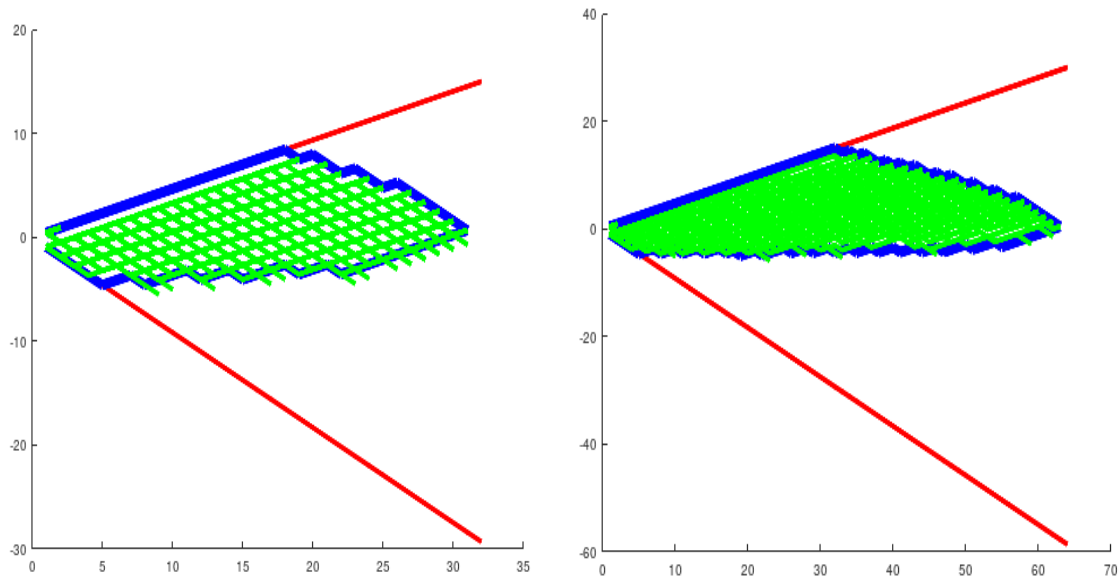
(a) $N = 32$, $E_1(\min\{H, n\}) = 10.53$ y (b) $N = 64$, $E_1(\min\{H, n\}) = 18.72$ y
 97.91 % decisiones correctas. 99.26 % decisiones correctas.

Figura 4.22: Observaciones de Bernoulli bajo $\theta_1 = 0.8$ (vs $\theta_0 = 0.5$) y tasa de fallos $p_H = 0.001$.



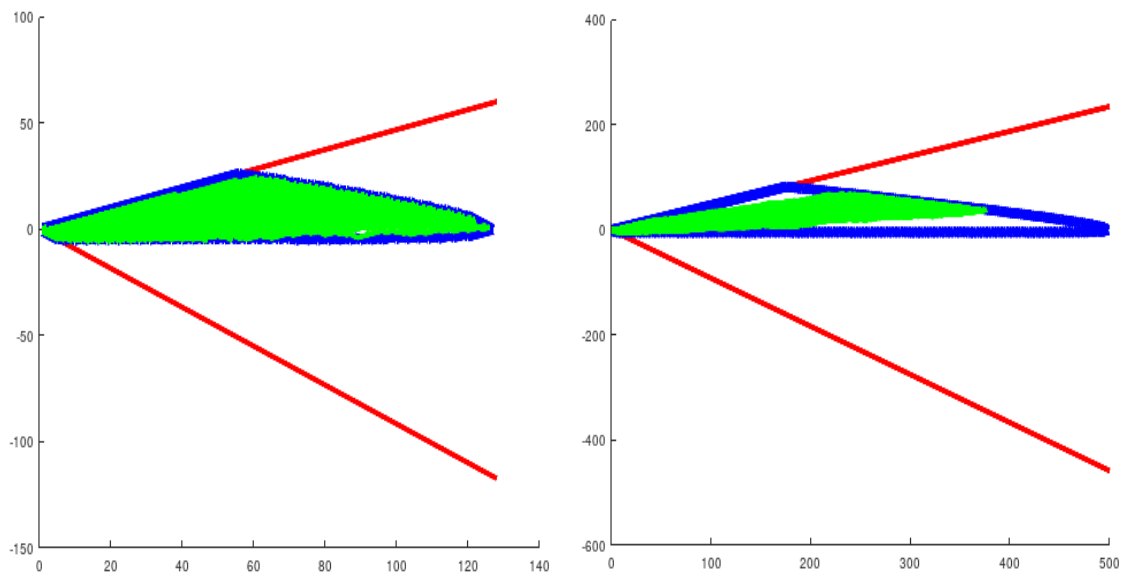
(a) $N = 128$, $E_1(\min\{H, n\}) = 34.13$ y (b) $N = 500$, $E_1(\min\{H, n\}) = 106.97$ y
 99.62 % decisiones correctas. 99.7 % decisiones correctas.

Figura 4.23: Observaciones de Bernoulli bajo $\theta_1 = 0.8$ (vs $\theta_0 = 0.5$) y tasa de fallos $p_H = 0.001$.



(a) $N = 32$, $E_1(\min\{H, n\}) = 10.58$ y (b) $N = 64$, $E_1(\min\{H, n\}) = 18.68$ y
 97.83 % decisiones correctas. 99.33 % decisiones correctas.

Figura 4.24: Observaciones de Bernoulli bajo $\theta_1 = 0.8$ (vs $\theta_0 = 0.5$) y tasa de fallos $p_H = 0.00001$.



(a) $N = 128$, $E_1(\min\{H, n\}) = 33.34$ y (b) $N = 500$, $E_1(\min\{H, n\}) = 108.29$ y
 99.78 % decisiones correctas. 99.61 % decisiones correctas.

Figura 4.25: Observaciones de Bernoulli bajo $\theta_1 = 0.8$ (vs $\theta_0 = 0.5$) y tasa de fallos $p_H = 0.00001$.

Capítulo 5

Conclusiones

Las conclusiones que podemos extraer de esta investigación y sobre todo de esta simulación por computadora para la obtención de la región de continuación son varias:

- Hemos obtenido una teoría más general y original sobre las pruebas secuenciales bajo el enfoque de un horizonte aleatorio, cuya distribución es independiente de la distribución de las observaciones de un proceso estocástico a tiempo discreto. Esto se puede ver cuando la tasa de fallos tiende a cero, $p_n \rightarrow 0$, con lo que tenemos $H \equiv \infty$.
- Falta más trabajo en cuanto al análisis de la constante K que aparece en la regla de decisión sobre uno de los parámetros de las hipótesis, debido a que no queda claro qué valor $K > 0$ pueda minimizar ambos errores cuando el horizonte aleatorio interrumpe un experimento secuencial y la razón de probabilidades queda dentro de la región de continuación. Durante las corridas del programa observamos que si $K \rightarrow A$, se mejora la decisión para error tipo II, mientras que si $K \rightarrow B$, (A, B) es nuestra región de continuación, por lo que falta verificar esto de manera numérica.
- Con la teoría desarrollada hemos determinado la estructura de las pruebas óptimas para las pruebas de Bayes para dos hipótesis compuestas, así como para procesos de Markov a tiempo discreto. Sólo hemos mostrado una estructura general de estos dos enfoques sin agotar el tema.
- En cuanto a los resultados computacionales cuando fijamos el intervalo de continuación (A, B) , hemos concluido que conforme los parámetros hipotetizados se encuentren relativamente alejados, nos proporciona una mejor decisión tanto para H_1 como para H_0 (hicimos más corridas pero no es posible poner todos los resultados en esta investigación). Con el valor $p_H = 0.00001$, equivalentemente $E(H) = 100000$, obtuvimos el 100 % de la decisión correcta, por lo que a partir de esta tasa de fallos o más pequeña tenemos la teoría clásica sin pasar al límite cuando $p_H \rightarrow 0$. Otra observación que obtuvimos de esta simulación computacional fue que mientras los parámetros hipotetizados estén relativamente cercanos, las decisiones correctas disminuyen para ambos casos; mientras que si están alejadas aumentarán. Desde luego, que sin importar qué tan cercanos estén los parámetros, si $p_H \rightarrow 0$ el porcentaje de una buena decisión tiende a 100 %.

- Para el problema inverso, es decir, para determinar la región de continuación, probamos con varios valores para λ y el costo por observación $c > 0$, siempre y cuando cumpliera la condición $\lambda > \frac{c}{\theta_1 - \theta_0}$. Al variar c vimos que la región de continuación se modificaba, por lo que consideramos estudiar más a fondo estos cambios en c como en [Cruz-Suárez et al. 2014](#). Queda abierto y con posibilidad de aplicar el método Montecarlo para la constante K , que es el punto para decidir sobre alguna de las decisiones cuando el horizonte detiene nuestro experimento secuencial y la razón de probabilidades z_n se encuentra en la región de continuación (A_n, B_n) . La idea es encontrar el K óptimo de tal manera que ambos errores (tipo I y II) se minimicen.
- Como otro trabajo futuro, vemos la posibilidad de aplicar el método Montecarlo para determinar la región de continuación para observaciones de Poisson y observaciones normales u otra distribución continua, pero discretizando los valores de la distribución, debido a las mismas restricciones computacionales.
- Gracias a este análisis numérico también vemos la posibilidad de realizar montecarlo para cuando el horizonte aleatorio posee otra distribución diferente a la geométrica.

Apéndice A

Demostraciones

A.1. Demostración del Teorema 2.1

Para probar (2.10), recordemos que $s_n^\psi = t_n^\psi \psi_n$ y $h_n^\psi = \psi_n P(H > n) + P(H = n)$, entonces

$$s_n^\psi P(H > n) + t_n^\psi P(H = n) = t_n^\psi h_n^\psi.$$

Así vemos que de (1.5) y (1.6) en la definición 1.2 tenemos que

$$\begin{aligned} \lambda\alpha(\psi, \phi) + \beta(\psi, \phi) &= \sum_{n=0}^{\infty} \lambda E_0 \phi_n t_n^\psi h_n^\psi + \sum_{n=0}^{\infty} E_1 (1 - \phi_n) t_n^\psi h_n^\psi \\ &= \sum_{n=0}^{\infty} \int t_n^\psi h_n^\psi (\lambda f_0^n \phi_n + f_1^n (1 - \phi_n)) d\mu^n. \end{aligned} \quad (\text{A.1})$$

Como $\lambda f_0^n \geq 0$, $f_1^n \geq 0$ y $0 \leq \phi_n \leq 1$, entonces podemos aplicar el resultado del Lema 2.2 a $F_1 = \lambda f_0^n$ y $F_2 = f_1^n$; por lo tanto, tenemos de manera inmediata que

$$\lambda\alpha(\psi, \phi) + \beta(\psi, \phi) \geq \sum_{n=0}^{\infty} \int t_n^\psi h_n^\psi \min\{\lambda f_0^n, f_1^n\} d\mu^n,$$

con igualdad si y sólo si

$$\phi_n \simeq I_{\{\lambda f_0^n \leq f_1^n\}} \quad \mu^n\text{-c.t.p. en } S_n^\psi \quad (\text{A.2})$$

para toda $n \geq 0$.

A.2. Demostración del Lema 2.3

Observemos que de la definición del funcional B_k^ψ en (2.19), probar (2.20) es equivalente a demostrar

$$\begin{aligned} &\int t_k^\psi h_k^\psi (ckf^k + l_k) d\mu^k + P(H \geq k+1) \int t_{k+1}^\psi (c(k+1)f^{k+1} + v) d\mu^{k+1} \\ &\geq P(H \geq k) \int t_k^\psi (ckf^k + D_k v) d\mu^k. \end{aligned} \quad (\text{A.3})$$

Por hipótesis tenemos que la función v es no negativa y μ^{k+1} -integrable; así, como las funciones que intervienen en (A.3) también lo son. Entonces factorizando a t_n^ψ y aplicando el Teorema de Fubini a la parte izquierda de (A.3) (haciendo $d\mu^{k+1} = d\mu^k d\mu(x_{k+1})$), tenemos que es igual a

$$\begin{aligned}
& \int t_k^\psi \left[h_k^\psi (ckf^k + l_k) \right. \\
& \quad \left. + (1 - \psi_k)P(H \geq k+1) \left(\int (c(k+1)f^{k+1} + v) d\mu(x_{k+1}) \right) \right] d\mu^k \\
&= \int t_k^\psi \left[(\psi_k P(H > k) + P(H = k))(ckf^k + l_k) \right. \\
& \quad \left. + (1 - \psi_k)P(H \geq k+1)(c(k+1)f^k + J_k v) \right] d\mu^k \\
&= P(H \geq k) \int t_k^\psi \left[(\psi_k(1 - p_k) + p_k)(ckf^k + l_k) \right. \\
& \quad \left. + (1 - \psi_k)(1 - p_k)(c(k+1)f^k + J_k v) \right] d\mu^k \\
&= P(H \geq k) \int t_k^\psi \left[(1 - p_k)(\psi_k(ckf^k + l_k) \right. \\
& \quad \left. + (1 - \psi_k)(c(k+1)f^k + J_k v)) + p_k(ckf^k + l_k) \right] d\mu^k \\
&= P(H \geq k) \int t_k^\psi \left[(1 - p_k)(ckf^k + (\psi_k l_k + (1 - \psi_k)(cf^k + J_k v))) \right. \\
& \quad \left. + p_k(ckf^k + l_k) \right] d\mu^k \\
&= P(H \geq k) \int t_k^\psi \left[ckf^k + p_k l_k + (1 - p_k)(\psi_k l_k + (1 - \psi_k)(cf^k + J_k v)) \right] d\mu^k. \quad (\text{A.4})
\end{aligned}$$

Lo anterior es cierto toda vez que al factorizar $P(H \geq k)$ (por hipótesis sabemos que $P(H \geq k+1) > 0$, luego también $P(H \geq k) > 0$) tenemos que

$$\begin{aligned}
\psi_k P(H > k) + P(H = k) &= P(H \geq k) \left(\psi_k \frac{P(H \geq k) - P(H = k)}{P(H \geq k)} + p_k \right) \\
&= P(H \geq k)(\psi_k(1 - p_k) + p_k)
\end{aligned}$$

y

$$\frac{P(H \geq k+1)}{P(H \geq k)} = \frac{P(H \geq k) - P(H = k)}{P(H \geq k)} = 1 - p_k.$$

Entonces aplicando el Lema 2.2 de Novikov (2009), vemos que (A.4) es mayor o igual

$$P(H \geq k) \int t_k^\psi \left[ckf^k + p_k l_k + (1 - p_k) \min\{l_k, cf^k + J_k v\} \right] d\mu^k \quad (\text{A.5})$$

$$= P(H \geq k) \int t_k^\psi (ckf^k + D_kv) d\mu^k,$$

lo cual coincide con la parte derecha de (A.3).

Tenemos además, por el mismo Lema 2.2, que hay igualdad entre (A.4) y (A.5) (que por transitividad corresponde a la igualdad en (A.3)) si y sólo si (2.22) se cumple μ^k -c.t.p. en T_k^ψ .

A.3. Demostración del Lema 2.4

Sea ψ cualquier regla de paro, tal que $E(\tau_\psi \wedge H) < \infty$; y sean λ y c dos números positivos cualesquiera. Entonces $\mathcal{L}(\psi) = \mathcal{L}(\psi; c, \lambda) < \infty$, y en virtud de (2.17) y (2.18),

$$\mathcal{L}(\psi) - \mathcal{L}_N(\psi) = \sum_{n=N}^{\infty} \int t_n^\psi h_n^\psi (cnf^n + l_n) d\mu^n - P(H \geq N) \int t_N^\psi (cNf^N + l_N) d\mu^N. \quad (\text{A.6})$$

Es inmediato ver que el primer sumando en (A.6) tiende a cero como $N \rightarrow \infty$, debido a que es la cola de una serie convergente.

Además, por el hecho de ser $E(\tau_\psi \wedge H) = \sum_{n=0}^{\infty} nP(\tau_\psi \wedge H = n) < \infty$ y usando la independencia de H con τ_ψ , tenemos

$$P(H \geq N) \int cNt_N^\psi f^N d\mu^N = cNP(H \geq N)P(\tau_\psi \geq N) = cNP(\tau_\psi \wedge H \geq N) \rightarrow 0$$

como $N \rightarrow \infty$.

Entonces ahora podemos ver fácilmente que, bajo el supuesto $E(\tau_\psi \wedge H) < \infty$, tenemos que $\mathcal{L}_N(\psi) \rightarrow \mathcal{L}(\psi)$, como $N \rightarrow \infty$, si y sólo si

$$P(H \geq N) \int t_N^\psi l_N d\mu^N \rightarrow 0,$$

como $N \rightarrow \infty$.

Finalmente supongamos que $E(\tau_\psi \wedge H) = \infty$, pero $P(\tau_\psi \wedge H < \infty) = 1$. Esto implica que

$$\mathcal{L}(\psi) = \sum_{n=1}^{\infty} \int t_n^\psi h_n^\psi (cnf_0^n + l_n) d\mu^n = \infty;$$

luego,

$$\mathcal{L}_N(\psi) \geq \sum_{n=1}^{N-1} \int t_n^\psi h_n^\psi (cnf_0^n + l_n) d\mu^n \rightarrow \infty, \quad N \rightarrow \infty.$$

A.4. Demostración del Corolario 2.3

Si ψ es tal que $P(\tau_\psi \wedge H < \infty) = 1$ entonces, por la independencia de τ_ψ y H ,

$$P(\tau_\psi \wedge H \geq n) = P(\tau_\psi \geq n, H \geq n) = P(\tau_\psi \geq n)P(H \geq n) \rightarrow 0, \quad (\text{A.7})$$

como $n \rightarrow \infty$.

Ahora si $P(H < \infty) = 1$ entonces (2.28) se sigue del Corolario 2.1.

Si $P(H < \infty) < 1$, entonces $\lim_{n \rightarrow \infty} P(H \geq n) = P(H = \infty) > 0$, luego podemos ver de (A.7) que $P(\tau_\psi \geq n) \rightarrow 0$, como $n \rightarrow \infty$, i.e. $P(\tau_\psi < \infty) = 1$. Ahora consideremos dos casos.

- a) $E\tau_\psi < \infty$. Se sigue del Lema 3.2 de Novikov (2009) y por el Corolario 3.2 también de Novikov (2009) que

$$\int t_n^\psi \min\{\lambda f_0^n, f_1^n\} \rightarrow 0, \text{ como } n \rightarrow \infty,$$

entonces, en virtud del Lema 2.4, concluimos que (2.28) se cumple.

- b) $E\tau_\psi = \infty$. En este caso,

$$E(\tau_\psi \wedge H) = \sum_{n=1}^{\infty} P(\tau_\psi \wedge H \geq n) = \sum_{n=1}^{\infty} P(\tau_\psi \geq n)P(H \geq n) = \infty$$

(porque $\lim_{n \rightarrow \infty} P(H \geq n) > 0$ y $E\tau_\psi = \sum_{n=1}^{\infty} P(\tau_\psi \geq n) = \infty$), y (2.28) se sigue de la segunda parte del Lema 2.4.

A.5. Demostración del Lema 2.5

Para demostrar este hecho, procedamos por inducción hacia atrás. Sea $n = N$, entonces

$$\begin{aligned} V_N^{N+1} = D_N V_{N+1}^{N+1} &= l_N p_N + (1 - p_N) \min\{l_N, c f^N + J_N l_{N+1}\} \\ &\leq l_N p_N + (1 - p_N) l_N = l_N = V_N^N. \end{aligned}$$

Supongamos que (2.29) se satisface para algún $0 \leq k+1 < N$. Entonces para $n = k$ tenemos

$$\begin{aligned} V_k^{N+1} = D_k V_{k+1}^{N+1} &= p_k l_k + (1 - p_k) \min\{l_k, c f^k + J_k V_{k+1}^{N+1}\} \\ &= p_k l_k + (1 - p_k) \min\{l_k, c f^k + \int V_{k+1}^{N+1} d\mu(x_{k+1})\} \\ &\leq p_k l_k + (1 - p_k) \min\{l_k, c f^k + \int V_{k+1}^N d\mu(x_{k+1})\} \\ &= p_k l_k + (1 - p_k) \min\{l_k, c f^k + J_k V_{k+1}^N\} \\ &= D_k V_{k+1}^N = V_k^N. \end{aligned}$$

Por lo tanto, vemos que también se satisface para k .

A.6. Demostración del Lema 3.1

Procedamos por inducción hacia atrás y supongamos que $f^n > 0$ para toda n . Si $k = N$ entonces

$$f^N \rho_N^N(z_N^0, z_N^1) = f^N g_\lambda(z_N^0, z_N^1) = f^N \min\{\lambda z_N^0, z_N^1\} = \min\{f_0^N \lambda, f_1^N\} = l_N = V_N^N.$$

Ahora supongamos que para un valor $k + 1$, con $0 \leq k + 1 < N$, se cumple la igualdad

$$f^{k+1} \rho_{k+1}^N(z_{k+1}^0, z_{k+1}^1) = V_{k+1}^N.$$

Entonces, para k tenemos que

$$\begin{aligned} f^k \rho_k^N(z_k^0, z_k^1) &= f^k g_\lambda(z_k^0, z_k^1) p_k + (1 - p_k) \min\{f^k g_\lambda(z_k^0, z_k^1), c f^k + f^k I_k \rho_{k+1}^N(z_k^0, z_k^1)\} \\ &= f^k g_\lambda(z_k^0, z_k^1) p_k + (1 - p_k) \min\left\{f^k g_\lambda(z_k^0, z_k^1), c f^k + I_k V_{k+1}^N\right\} \\ &= l_k p_k + (1 - p_k) \min\{l_k, c f^k + J_k V_{k+1}^N\} \\ &= D_k V_{k+1}^N = V_k^N. \end{aligned}$$

A.7. Demostración del Lema 3.2

Procedamos por inducción hacia atrás. Si $k = N$ entonces

$$\begin{aligned} \rho_N^{N+1}(z^0, z^1) &= g_\lambda(z^0, z^1) p_N + (1 - p_N) \min\{g_\lambda(z^0, z^1), c + I_N \rho_{N+1}^{N+1}(z^0, z^1)\} \\ &\leq g_\lambda(z^0, z^1) = \rho_N^N(z^0, z^1), \end{aligned}$$

para toda $z^0, z^1 \geq 0$ y $c > 0$.

Ahora supongamos que se cumple $\rho_k^N(z^0, z^1) \geq \rho_k^{N+1}(z^0, z^1)$ (y con ello también se satisface $I_{k-1} \rho_k^N(z^0, z^1) \geq I_{k-1} \rho_k^{N+1}(z^0, z^1)$, porque sólo se toma la esperanza bajo H en la desigualdad) para alguna k con $0 \leq k < N$, para toda $z \in \mathbb{R}^+$. Entonces

$$\begin{aligned} \rho_{k-1}^N(z^0, z^1) &= g_\lambda(z^0, z^1) p_{k-1} + (1 - p_{k-1}) \min\{g_\lambda(z^0, z^1), c + I_{k-1} \rho_k^N(z^0, z^1)\} \\ &\geq g_\lambda(z^0, z^1) p_{k-1} + (1 - p_{k-1}) \min\{g_\lambda(z^0, z^1), c + I_{k-1} \rho_k^{N+1}(z^0, z^1)\} \\ &= \rho_{k-1}^{N+1}(z^0, z^1), \end{aligned}$$

para toda $z^0, z^1 \in \mathbb{R}^+$. Por lo tanto, también se satisface para $k - 1$.

A.8. Demostración del Lema 3.3

Primero recordemos que si ν es una función cóncava, entonces por la desigualdad de Jensen $I_n \nu = E_0 \nu$ también es cóncava.

Iniciemos la demostración del inciso (i) por inducción hacia atrás tanto para ρ_N^N como para $I_{N-1}\rho_N^N$. Supongamos que $N \geq 0$ es fijo, entonces para un entero no negativo n tal que $n = N$ tenemos que $\rho_N^N(z) = g_\lambda(z) = \min\{\lambda, z\}$ cumple inmediatamente con las propiedades de $\mathcal{K}$, es decir, es continua, no decreciente, cóncava (por ser el mínimo de dos funciones cóncavas) en $\mathbb{R}^+$, $\rho_N^N(0) = 0$, $\lim_{z \rightarrow \infty} \rho_N^N(z) = \lambda$. Además, se cumple $\rho_N^N(z) \leq \min\{\lambda, z\}$ para toda $z \in \mathbb{R}^+$.

Por otro lado, y debido a ρ_N^N , la función $I_{N-1}\rho_N^N$ es cóncava, asimismo continua y no decreciente en $\mathbb{R}^+$. Es inmediato ver que $I_{N-1}\rho_N^N(0) = 0$. Si aplicamos el teorema de convergencia monótona y tomando en cuenta la condición 3.33, tenemos que $\lim_{z \rightarrow \infty} I_{N-1}\rho_N^N(z) = \lambda$. Finalmente, es fácil ver que $I_{N-1}\rho_N^N(z) \leq \min\{\lambda, z\}$.

Supongamos ahora que para alguna $k < N$, tenemos que $\rho_k^N, I_{k-1}\rho_k^N \in \mathcal{K}$. Entonces para $k-1$, tenemos que ρ_{k-1}^N (ver definición 3.36) es también cóncava por ser suma de funciones cóncavas, por lo que es continua y de la misma definición tenemos también que es no decreciente en $\mathbb{R}^+$. También es fácil ver que $\rho_{k-1}^N(0) = 0$ y

$$\begin{aligned} \lim_{z \rightarrow \infty} \rho_{k-1}^N(z) &= \lim_{z \rightarrow \infty} \left(g_\lambda(z)p_{k-1} + (1 - p_{k-1}) \min\{g_\lambda(z), c + I_{k-1}\rho_k^N(z)\} \right) \\ &= \lambda p_{k-1} + (1 - p_{k-1}) \min\{\lambda, c + \lim_{z \rightarrow \infty} I_{k-1}\rho_k^N(z)\} \\ &= \lambda p_{k-1} + (1 - p_{k-1}) \min\{\lambda, c + \lambda\} = \lambda, \quad \text{por que } c > 0. \end{aligned}$$

Finalmente tenemos de la definición que $\rho_{k-1}^N(z) \leq \min\{\lambda, z\}$ para toda $z \in \mathbb{R}^+$, por lo que también $\rho_{k-1}^N \in \mathcal{K}$.

Por otro lado, como ρ_{k-1}^N es cóncava, entonces $I_{k-2}\rho_{k-1}^N$ también es cóncava, luego también continua y por la misma definición de $I_{k-2}\rho_{k-1}^N$, también vemos que es no decreciente. Ahora por el teorema de la convergencia monótona y la condición 3.33 tenemos que $\lim_{z \rightarrow 0} I_{k-2}\rho_{k-1}^N(z) = 0$ y $\lim_{z \rightarrow \infty} I_{k-2}\rho_{k-1}^N(z) = \lambda$ y de la misma definición vemos que $I_{k-2}\rho_{k-1}^N(z) \leq \min\{\lambda, z\}$ para toda $z \in \mathbb{R}^+$. Por lo tanto, para $k-1$ también $I_{k-2}\rho_{k-1}^N \in \mathcal{K}$.

La demostración del punto (ii), se sigue de manera análoga de la demostración del Lema 3.2.

Para demostrar el punto (iii) procedamos nuevamente por inducción hacia atrás. Sea $n = N$, entonces la función límite $\rho_N(z) = \lim_{N \rightarrow \infty} \rho_N^N(z) = \lim_{N \rightarrow \infty} g_\lambda(z) = g_\lambda(z)$, la cual por el punto (i) cumple con $\rho_N \in \mathcal{K}$. Podemos ver que $I_{N-1}\rho_N(z) = g_\lambda(z)$, por lo tanto $I_{N-1}\rho_N \in \mathcal{K}$.

Ahora supongamos que para algún k , se tiene $\rho_k \in \mathcal{K}$ y $I_{k-1}\rho_k \in \mathcal{K}$. Demostremos primero que para $k-1$, $\rho_{k-1} \in \mathcal{K}$. La función límite $\rho_{k-1}(z) = \lim_{N \rightarrow \infty} \rho_{k-1}^N(z)$ también es una función cóncava por ser límite de funciones cóncavas, toda vez que también $I_{k-1}\rho_k(z)$ es cóncava (por hipótesis) y por consiguiente es continua. Asimismo, es una función no decreciente por ser límite de funciones no decrecientes. También tenemos

$\rho_{k-1}(0) = \lim_{N \rightarrow \infty} \rho_{k-1}^N(0) = 0$. Ahora bien,

$$\begin{aligned} \lim_{z \rightarrow \infty} \rho_{k-1}(z) &= \lim_{z \rightarrow \infty} \left(g_\lambda(z) p_{k-1} + (1 - p_{k-1}) \min\{g_\lambda(z), c + I_{k-1}\rho_k(z)\} \right) \\ &= \lambda p_{k-1} + (1 - p_{k-1}) \min\{\lambda, c + \lim_{z \rightarrow \infty} I_{k-1}\rho_k^N(z)\} \\ &= \lambda p_{k-1} + (1 - p_{k-1}) \min\{\lambda, c + \lambda\} = \lambda, \quad \text{por que } c > 0. \end{aligned}$$

Finalmente y por los incisos (i) y (ii), tenemos que $\rho_{k-1}(z) \leq \rho_{k-1}^N(z) \leq \min\{\lambda, z\}$ para toda $z \in \mathbb{R}^+$. Por lo tanto $\rho_{k-1} \in \mathcal{K}$.

Ahora demostremos que $I_{k-2}\rho_{k-1} \in \mathcal{K}$. Como $\rho_{k-1} \in \mathcal{K}$, tenemos que $I_{k-2}\rho_{k-1}$ es cóncava, continua y no decreciente por ser límite de funciones cóncavas, y no decrecientes. La continuidad la obtenemos de la concavidad. Asimismo, vemos que $I_{k-2}\rho_{k-1}(0) = 0$ y $\lim_{z \rightarrow \infty} I_{k-2}\rho_{k-1}(z) = \lambda$ por la condición 3.33 y el teorema de convergencia monótona. Para el último punto también es sencillo ver que la función límite $I_{n-2}\rho_{k-1}(z) \leq \min\{\lambda, z\}$ para toda $z \in \mathbb{R}^+$, por lo que $I_{k-2}\rho_{k-1} \in \mathcal{K}$.

El inciso (iv) sólo es la definición de $\rho_n(z)$ aplicando el límite como $N \rightarrow \infty$ en $\rho_n^N(z)$, para toda $n \geq 0$ y toda $z \in \mathbb{R}^+$.

A.9. Demostración del Lema 3.6

Dada cualquier $n \geq 0$ definamos a $u_n, v_n : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ como $u_n(z) = z - I_n\rho_{n+1}(z)$ y $v_n(z) = \lambda - I_n\rho_{n+1}(z)$, respectivamente, para toda $z \in \mathbb{R}^+$. Entonces, usando el hecho que, debido al Lema 3.3, $I_n\rho_{n+1} \in \mathcal{K}$, no es difícil ver que las siguientes propiedades tienen lugar.

- (i) u_n y v_n son continuas, no negativas y convexas. La continuidad la obtenemos debido a que u_n, v_n están definidas como suma de funciones continuas; son no negativas toda vez que $I_n\rho_{n+1}(z) \leq g_\lambda(z)$ para toda $z \geq 0$. La convexidad la obtenemos porque z es lineal y $I_n\rho_{n+1}$ es cóncava.
- (ii) $u_n(0) = 0$, $\lim_{z \rightarrow \infty} u_n(z) = \infty$, $v_n(0) = \lambda$, $\lim_{z \rightarrow \infty} v_n(z) = 0$. Se sigue por la propiedad (b) de $\mathcal{K}$.
- (iii) Para toda $x, y \in \mathbb{R}^+$ tal que $x < y$
 - (a) $u_n(x) < u_n(y)$, siempre que $u_n(x) > 0$,
 - (b) $v_n(x) > v_n(y)$, siempre que $v_n(y) > 0$.
- (iv) $u_n(\lambda) = v_n(\lambda) = \lambda - I_n\rho_{n+1}(\lambda)$. Esta expresión es estrictamente mayor que c si y sólo si (3.43) se cumple.

Enseguida verifiquemos la propiedad (iii) por contradicción en ambos casos.

Probemos primero que dados $x, y \in \mathbb{R}^+$ con $x < y$ y $u_n(x) > 0$, entonces $u_n(x) < u_n(y)$. Para lo anterior supongamos lo contrario, es decir que $u_n(x) \geq u_n(y)$. Sabemos que existe un $\alpha \in (0, 1)$, tal que $x = \alpha 0 + (1 - \alpha)y = (1 - \alpha)y$. Ahora utilizando la convexidad de u_n tenemos que

$$u_n(y) \leq u_n(x) = u_n((1 - \alpha)y) \leq (1 - \alpha)u_n(y) < u_n(y) \quad \forall \alpha \in (0, 1),$$

lo cual es una contradicción. Por lo tanto $u_n(x) < u_n(y)$.

Finalmente probemos que si $x, y \in \mathbb{R}^+$ con $x < y$ y $v_n(y) > 0$, entonces $v_n(x) > v_n(y)$. Supongamos que $v_n(x) \leq v_n(y)$, entonces por la continuidad de v_n y por el hecho de que $v_n(\infty) = 0$ (propiedades (i) y (ii), respectivamente), podemos tomar un $z > y$, de tal forma que $0 < v_n(z) < v_n(y)/2$. También sabemos que para toda $t \in (x, z)$ existe un $\alpha \in (0, 1)$, tal que $t = \alpha x + (1 - \alpha)z$. Entonces para y existe un α tal que $y = \alpha x + (1 - \alpha)z$. Ahora, por la convexidad de v_n , tenemos que

$$\begin{aligned} v_n(y) &= v_n(\alpha x + (1 - \alpha)z) \leq \alpha v_n(x) + (1 - \alpha)v_n(z) \\ &\leq \alpha v_n(y) + (1 - \alpha)\frac{v_n(y)}{2} \\ &= (1 + \alpha)\frac{v_n(y)}{2} < v_n(y) \quad \text{para toda } \alpha \in (0, 1). \end{aligned}$$

Con lo cual también llegamos a una contradicción. Por lo tanto, se debe cumplir que $v_n(x) > v_n(y)$.

Supongamos ahora que la desigualdad (3.43) se cumple. Entonces de (i)–(iv), vemos que para una $c > 0$, es obvio (aplicando el Teorema del Valor Intermedio) que la ecuación $u_n(z) = c$ tiene una solución única en $z = A_n < \lambda$, y que $u_n(z) = z - I_n \rho_{n+1}(z) \leq c$ si y sólo si $z \leq A_n$.

De la misma manera, la ecuación $v_n(z) = c$ tiene una única solución en $z = B_n$, con $B_n > \lambda$, y $v_n(z) = \lambda - I_n \rho_{n+1}(z) \leq c$ si y sólo si $z \geq B_n$. Debido a que el $\min\{u_n(z), v_n(z)\} = g_\lambda(z) - I_n \rho_{n+1}(z)$, tenemos que la primera afirmación del Lema 3.6 se cumple.

Si (3.43) no se cumple, esto es, si $\lambda \leq c + I_n \rho_{n+1}(z)$, tenemos de (i)–(iv) que $u_n(z) \leq c$ para toda $z \leq \lambda$ y $v_n(z) \leq c$ para toda $z \geq \lambda$, entonces $\min\{u_n(z), v_n(z)\} = g_\lambda(z) - I_n \rho_{n+1}(z) \leq c$ para toda $z \geq 0$, lo cual demuestra que la última afirmación del Lemma.

A.10. Demostración del Teorema 3.4

Para proceder a la demostración de este teorema necesitamos algunos lemas que enseguida enunciaremos y probaremos.

Tomando en cuenta el Lema 3.5, podemos definir de manera natural cuando $c = 0$ a

$$\rho_0(z; 0, \lambda) = \inf_{\psi, \phi} \{\lambda \alpha(\psi, \phi) + z \beta(\psi, \phi)\}, \quad (\text{A.8})$$

para toda $z > 0$ y $\lambda > 0$.

Lema A.1. Para toda $z > 0$ y $\lambda > 0$

$$\rho_0(z; 0, \lambda) = \sum_{n=0}^{\infty} (\lambda P_0(z \mathcal{Z}_n > \lambda) + z P_1(z \mathcal{Z}_n \leq \lambda)) P(H = n). \quad (\text{A.9})$$

Demostración. Fijemos a un entero no negativo cualquiera N , y haciendo uso del Teorema 2.2, se sigue que

$$\inf_{\psi \in \mathcal{D}^N, \phi} \{ \lambda \alpha(\psi, \phi) + z \beta(\psi, \phi) \} = \rho_0^N(z; 0, \lambda), \quad (\text{A.10})$$

donde, debido a (3.58), $\rho_n^N(z) = D^{N-n} g_\lambda(z)$, $0 \leq n \leq N$, siendo D definida en (3.57) (tomando precisamente a $c = 0$).

Se sigue del Teorema 2.2 que el ínfimo en (A.10) es alcanzado, en particular, si ψ es tal que $\psi_n = I_{\{g_\lambda(Z_n) < I_n \rho_{n+1}^N(Z_n)\}}$, $n = 0, 1, \dots, N-1$ y $\psi_N \equiv 1$. Es fácil notar que la parte (i) del Lema 3.3 también aplica para $c = 0$, entonces se sigue que $I_n \rho_{n+1}^N(z) \leq g_\lambda(z)$ para toda $z \geq 0$. Por lo tanto, una regla de paro óptima es $\psi^N \in \mathcal{D}^N$ si

$$\psi_n^N \equiv 0 \quad \text{para} \quad 0 \leq n \leq N-1, \quad \text{y} \quad \psi_N^N \equiv 1, \quad (\text{A.11})$$

con

$$\begin{aligned} \inf_{\phi} \{ \lambda \alpha(\psi^N, \phi) + z \beta(\psi^N, \phi) \} &= \rho_0^N(z; 0, \lambda) \\ &= \sum_{n=0}^{N-1} (\lambda P_0(z \mathcal{Z}_n > \lambda) + z P_1(z \mathcal{Z}_n \leq \lambda)) P(H = n) \\ &\quad + (\lambda P_0(z \mathcal{Z}_N > \lambda) + z P_1(z \mathcal{Z}_N \leq \lambda)) P(H \geq N) \end{aligned} \quad (\text{A.12})$$

Es obvio de (A.12) que $\rho_0^N(z; 0, \lambda)$ converge, como $N \rightarrow \infty$, a la parte derecha de (A.9), i.e. $\rho_0(z; 0, \lambda)$.

También es inmediato ver que la demostración del Lema 2.4 (ver (A.6)) que

$$\mathcal{L}_N(\psi; 0, \lambda) \rightarrow \mathcal{L}(\psi; 0, \lambda)$$

para toda regla de paro ψ , entonces se sigue que $\mathcal{L}(\psi; 0, \lambda) \geq \rho(1; 0, \lambda)$ (cf. (3.45)), y

$$\inf_{\phi} \{ \lambda \alpha(\psi, \phi) + z \beta(\psi, \phi) \} \geq \rho(z; 0, \lambda). \quad (\text{A.13})$$

para toda ψ . Ahora bien si ψ es tal que $\psi_n \equiv 0$ para toda $n \geq 0$, se verifica fácilmente que el ínfimo en (A.13) coincide con la parte derecha de (A.9), entonces (A.9) se sigue de la definición de $\rho(z; 0, \lambda)$ en (A.8). $\square$

Notemos que el ínfimo en ambas ecuaciones (3.41) y (A.8) son cero cuando $\lambda = 0$ o $z = 0$. Entonces con base en lo anterior podemos definir a $\rho(0; c, \lambda)$ y $\rho(z; c, 0)$ como 0 para toda $z \geq 0$, $c \geq 0$ y $\lambda \geq 0$. Así hacemos a la función $\rho(z; c, \lambda)$ definida para toda $z \geq 0$, $c \geq 0$ y $\lambda \geq 0$.

El siguiente lema nos muestra que $\rho(z; c, \lambda)$ es continua en $c = 0$ cuando $z \geq 0$ y $\lambda \geq 0$ se mantienen fijas.

Lema A.2. *Supongamos que se satisface (3.53). Entonces para toda $z \geq 0$ y $\lambda \geq 0$,*

$$\lim_{c \rightarrow 0} \rho(z; c, \lambda) = \rho(z; 0, \lambda).$$

Demostración. Para toda $c > 0$

$$\begin{aligned}
0 &\leq \rho(z; c, \lambda) - \rho(z; 0, \lambda) \\
&\leq \inf_{\psi, \phi} \{c\mathcal{N}(\psi) + \lambda\alpha(\psi, \phi) + z\beta(\psi, \phi)\} - \rho(z; 0, \lambda) \\
&\leq \inf_{\phi} \{c\mathcal{N}(\psi^N) + \lambda\alpha(\psi^N, \phi) + z\beta(\psi^N, \phi)\} - \rho(z; 0, \lambda),
\end{aligned} \tag{A.14}$$

donde ψ^N está definida por (A.11). Obviamente el lado derecho de (A.14) es igual a

$$\begin{aligned}
&cE_0(H \wedge N) - \sum_{n=N}^{\infty} (\lambda P_0(z\mathcal{Z}_n > \lambda) + zP_1(z\mathcal{Z}_n \leq \lambda)) P(H = n) \\
&\quad + (\lambda P_0(z\mathcal{Z}_N > \lambda) + zP_1(z\mathcal{Z}_N \leq \lambda)) P(H \geq N) \\
&\leq cN + \lambda P_0(z\mathcal{Z}_N > \lambda) + zP_1(z\mathcal{Z}_N \leq \lambda).
\end{aligned} \tag{A.15}$$

Sea $\epsilon > 0$ y sea N lo suficientemente grande de tal forma que el segundo y tercer sumando en (A.15) no exceda ϵ (podemos hacer esto por el supuesto en (3.53)). Se sigue de (A.14) y (A.15) que

$$0 \leq \rho(z; c, \lambda) - \rho(z; 0, \lambda) \leq 3\epsilon, \quad \text{si } c < \epsilon/N.$$

□

Ahora sean A y B tales que $0 < A < B$. Probaremos que existe $c > 0$ y λ , $A \leq \lambda \leq B$, tales que

$$A = c + I\rho(A; c, \lambda) \quad \text{y} \quad \lambda = c + I\rho(B; c, \lambda), \tag{A.16}$$

donde $\rho(z; c, \lambda) (= \rho(z))$ es definida en (3.59).

Para toda $\lambda \geq A$ fija, definamos, a partir de la primera de las ecuaciones en (A.16), a $c = c(\lambda)$. Para demostrar que queda bien definida probemos las siguientes propiedades de la función $d_\lambda(c) = c + I\rho(A; c, \lambda)$.

Lema A.3. *Supongamos que se satisface (3.53). Entonces*

(i) $d_\lambda(c)$ es no negativa, estrictamente creciente y continua para $c \geq 0$.

(ii) $d_\lambda(0) < A$.

(iii) $\lim_{c \rightarrow \infty} d_\lambda(c) = \infty$.

Demostración. La parte (i) se sigue del Lema 3.5 y del Corolario 3.1. La parte (iii) es inmediata.

Probemos ahora la parte (ii).

Con base en cálculos tenemos que

$$I\rho(A; 0, \lambda) = \sum_{n=0}^{\infty} (\lambda P_0(A\mathcal{Z}_{n+1} > \lambda) + AP_1(A\mathcal{Z}_{n+1} \leq \lambda)) P(H = n). \tag{A.17}$$

Además, para toda $n \geq 1$

$$\lambda P_0(A\mathcal{Z}_n > \lambda) + AP_1(A\mathcal{Z}_n \leq \lambda) \leq A, \tag{A.18}$$

como

$$\begin{aligned}\lambda P_0(A\mathcal{Z}_n > \lambda) &= \int_{\{Af_1^n > \lambda f_0^n\}} \lambda f_0^n d\mu^n \\ &\leq \int_{\{Af_1^n > \lambda f_0^n\}} Af_1^n d\mu^n = AP_1(A\mathcal{Z}_n > \lambda).\end{aligned}$$

Se sigue de (A.17) y (A.18) que $d_\lambda(0) = I\rho(A; 0, \lambda) \leq A$, con igualdad si y sólo si hay igualdades en (A.18) para toda $n \geq 1$.

Pero, bajo la suposición de (3.53),

$$P_0(A\mathcal{Z}_n > \lambda) \rightarrow 0, \text{ y } P_1(A\mathcal{Z}_n \leq \lambda) \rightarrow 0, \text{ como } n \rightarrow \infty,$$

entonces hay desigualdad estricta en (A.18) para n grande. Por lo tanto, queda demostrado en punto (ii). $\square$

Ahora, para cualquier $\lambda \geq A$, $c = c(\lambda)$ se define de manera única por

$$c(\lambda) + I\rho(A; c(\lambda), \lambda) = A, \quad (\text{A.19})$$

y $c(\lambda) > 0$ es una función continua y estrictamente creciente de λ , en virtud del Teorema de la Función Implícita.

Demostremos que existe λ , $A \leq \lambda \leq B$ tal que

$$\lambda = c(\lambda) + I\rho(B; c(\lambda), \lambda) \quad (\text{A.20})$$

(cf. (A.16)). Para tal fin, definamos la función

$$G(\lambda) = \lambda - c(\lambda) - I\rho(B; c(\lambda), \lambda),$$

la cual es una función continua de λ , en virtud del Corolario 3.1. Ahora bien, si demostramos que $G(A) \leq 0$ y $G(B) \geq 0$, entonces la existencia de $\lambda \in [A, B]$ que satisface (A.20) queda garantizada en virtud del Teorema del Valor Intermedio.

En efecto, como $c(A) = A - I\rho(A; c(A), A)$, debido a (A.19), tenemos entonces el primer resultado

$$G(A) = A - c(A) - I\rho(B; c(A), A) = I\rho(A; c(A), A) - I\rho(B; c(A), A) \leq 0.$$

Ahora, usando (A.19) nuevamente, $c(B) = A - I\rho(A; c(B), B)$, con lo cual

$$\begin{aligned}G(B) &= B - c(B) - I\rho(B; c(B), B) \\ &= B - I\rho(B; c(B), B) - (A - I\rho(A; c(B), B)).\end{aligned} \quad (\text{A.21})$$

Observemos que $z - I\rho(z; c(B), B)$ es una función convexa no negativa de z e igual a cero en el valor $z = 0$ y no decreciente. Entonces a partir de (A.21) tenemos que $G(B) \geq 0$.

Hemos demostrado justamente que, para cualesquiera que sean A, B con $0 < A < B$, entonces existen $c > 0$ y $\lambda \in [A, B]$ tales que (A.16) se cumple.

Ahora si $1 \in [A, B]$, entonces la afirmación del Teorema 3.4 (con $K = \lambda$) se sigue del Teorema 3.3.

Apéndice B

Programas en GNU Octave

B.1. Código para estimar errores tipo I, II y $E_i(\min\{H, n\})$

Este y el siguiente código fueron implementados con uso del software [GNU Octave](#). Para su correcta ejecución se debe invocar el nombre de MonteCarlo en la terminal de GNU Octave.

```
#Método Montecarlo con 10000 réplicas para estimar los ←  
    errores tipo I y II,  
#tomando como dominio el parámetro p de la distribución ←  
    geométrica debida al  
#horizonte aleatorio, para probar H0 vs H1 (simples), ←  
    asimismo para estimar  
#el tamaño muestral esperado  $E_i[\min(N, H)]$   $i=0,1$ , de la ←  
    prueba bajo H0 y bajo H1  
  
# Opcional: Instalar y cargar paquetes necesarios ←  
    statisticsm io  
##pkg install -forge io  
##pkg install -forge statistics  
##pkg load io  
pkg load statistics  
tic  
clear all  
#Fijamos las cotas alpha y beta.  
a=.1;b=.1;  
A=log(b/(1-a));B=log((1-b)/a);  
#En caso de término por el horizonte H, rechazamos H1 si  $\ln ←  
    (z_n) \leq K$ .  
K=0;  
#Elegimos el tipo de observaciones y si el horizonte ←  
    intervendrá frecuentemente o no.  
DistF=input("Generar Bernoulli Prueba1:(1), Bernoulli ←
```

```

Prueba2: (2), Exponencial Prueba1: (3), Exponencial ←
Prueba2: (4) ←";
Conjuntop=input("Qué valores desea para p: (1) ←
0.001,0.002,...,0.01 ó (2) 0.01,0.02,...,0.1 ←");
#Se elige que el horizonte intervenga poco
if Conjuntop==1
    #Se elige que el horizonte intervenga menos E(H)=100 a ←
    1000
    p=0.001:.001:.01;
else
    #Se elige que el horizonte intervenga más E(H)=10 a 100
    p=0.01:0.01:0.1;
end
#Número total de réplicas
m=10000;
#Se eligen observaciones de Bernoulli en la prueba 1
if DistF==1
    texto='Caso discreto: Distribución Bernoulli Prueba 1';
    obs='Bernoulli';
    #Valores de H0 y H1, respectivamente
    t=[0.5 0.53];
    v1=log(t(2)*(1-t(1))/(t(1)*(1-t(2)))); v2=log((1-t(2))/(1-←
    t(1)));
    #Se eligen observaciones de Bernoulli en la prueba 2
elseif DistF==2
    texto='Caso discreto: Distribución Bernoulli Prueba 2';
    obs='Bernoulli';
    t=[0.5 0.8];
    v1=log(t(2)*(1-t(1))/(t(1)*(1-t(2)))); v2=log((1-t(2))←
    /(1-t(1)));
    #Se eligen observaciones de exponenciales en la prueba ←
    1
elseif DistF==3
    texto='Caso continuo: Distribución Exponencial Prueba 1 ←
    ';
    obs='Exponencial';
    t=[5 6];
    v1=1/t(1)-1/t(2); v2=log(t(1)/t(2));
    #Se eligen observaciones de exponenciales en la prueba ←
    2
elseif DistF==4
    texto='Caso continuo: Distribución Exponencial Prueba 2 ←
    ';
    obs='Exponencial';
    t=[5 11];

```

```

    v1=1/t(1)-1/t(2); v2=log(t(1)/t(2));
end
#Inicializamos variables para identificar decisiones por  $H_i \leftarrow$ 
    cuando el paro
#ocurre por horizonte o bien por la regla de paro.
HOD=zeros(2,length(p),m); H1D=zeros(2,length(p),m);
HOH=zeros(2,length(p),m); H1H=zeros(2,length(p),m);
#Hipótesis  $i=1,2$ , nula y alterna, respectivamente
for i=1:2
    #Dominio p
    for j=1:length(p)
        #m=10000 réplicas de una misma prueba
        for k=1:1:m
            #Generamos horizonte  $H \sim \text{Geo}(p(j))$  en cada réplica
            H=geornd(p(j)); r=0; fin=0; s=0;
            #Pruebas 1 ó 2 de Bernoulli
            if DistF==1 || DistF==2
                do
                    #Contador para el tiempo de paro (ya sea por  $N$  o  $\leftarrow$ 
                        por  $H$ )
                    ++r;
                    #Estadístico
                    s=v1*binornd(1,t(i))+v2+s;
                    #Si  $H$  no ha llegado
                    if H>r
                        #Aceptamos  $H_0$ 
                        if s<=A
                            # $N=0$  es terminar por regla de paro
                            N(i,j,k)=0; fin=1; HOD(i,j,k)=1;
                            #Rechazamos  $H_0$ 
                        elseif s>=B
                            N(i,j,k)=0; fin=1; H1D(i,j,k)=1;
                        end
                        #Si  $H$  interrumpe la prueba lo identificamos con  $N \leftarrow$ 
                            (i,j,k)=1
                    else
                        #Si fin=1 finaliza la m réplica
                        N(i,j,k)=1; fin=1;
                        #Aceptamos  $H_0$  y se registra en HOH.
                        if s<=K
                            HOH(i,j,k)=1;
                            #Aceptamos  $H_1$  y se registra en H1H
                        else H1H(i,j,k)=1;
                        end
                    end
                end
            end
        end
    end
end

```

```

    until(fin==1)
    #Registramos tiempo de paro del experimento
    n(i,j,k)=r;
    #Pruebas 1 ó 2 Exponencial análogas a lo anterior
elseif DistF==3 || DistF==4
    do
        ++r;
        s=v1*exprnd(t(i))+v2+s;
        if H>r
            if s<=A
                N(i,j,k)=0;fin=1;H0D(i,j,k)=1;
            elseif s>=B
                N(i,j,k)=0;fin=1;H1D(i,j,k)=1;
            end
        else
            N(i,j,k)=1;fin=1;
            if s<=K
                H0H(i,j,k)=1;
            else H1H(i,j,k)=1;
            end
        end
    until(fin==1)
    n(i,j,k)=r;
end
end
end
end
#Inicializamos variables para evitar errores
EN_0=zeros(1,length(p));
EN_1=zeros(1,length(p));
TPH0H=zeros(1,length(p));
TPH1H=zeros(1,length(p));
D1_H0=zeros(1,length(p));
D0_H1=zeros(1,length(p));
for i=1:length(p)
    #Calculamos promedios de paro bajo H0
    EN_0(i)=mean(n(1,i,:));
    #Calculamos promedios de paro bajo H1
    EN_1(i)=mean(n(2,i,:));
    #Contamos cuántas veces interviene H bajo H0
    TPH0H(i)=sum(N(1,i,:));
    #Contamos cuántas veces interviene H bajo H1
    TPH1H(i)=sum(N(2,i,:));
    #Estimación del error Tipo I
    D1_H0(i)=(sum(H1D(1,i,:)))/m+(sum(H1H(1,i,:)))/m;

```

```

#Estimación del error Tipo II
DO_H1(i)=(sum(HOD(2,i,:)))/m+(sum(HOH(2,i,:)))/m;
end

#Resultados gráficos de la estimación de los promedios del tamaño muestral
subplot(1,2,1)
plot(p,EN_0,'b','linewidth',6,p,EN_1,'k','linewidth',6)
set(gca,'FontSize',28)
title(['Estimación:E_0[min(H,\tau_\psi)], E_1[min(H,\tau_\psi)]'], 'FontName','Arial','FontSize',28);
axis([min(p) max(p) -0.1 1.3*max(max(EN_0),max(EN_1))]);
xlabel({'Tasa de fallos p', ['Interviene H en promedio', num2str(100*((mean(TPH0H)+mean(TPH1H))/m)/2,3), '%']}, 'FontName','Arial','FontSize',28);
ylabel(('Promedio muestral bajo H_0 y H_1'), 'FontSize',28);
#legend('E_0[min(n,H)]', 'E_1[min(n,H)]');
#Resultados gráficos de la estimación de las probabilidades de los errores tipo I y II
subplot(1,2,2)
y=ones(1,length(p)).*1;
plot(p,y,'r','linewidth',6,p,D1_H0,'b','linewidth',6,p,D0_H1,'k','linewidth',6)
set(gca,'FontSize',28)
title(['Estimación errores tipo I y II'], 'FontName','Arial','FontSize',28);
axis([min(p) max(p) -.01 3/2*max(max(D1_H0),max(D0_H1))]);
xlabel({'Tasa de fallos p', ['Interviene H en promedio', num2str(100*((mean(TPH0H)+mean(TPH1H))/m)/2,3), '%']}, 'FontName','Arial','FontSize',28);
ylabel(('Estimación errores tipo I y II'), 'FontSize',28);
#legend('Cota=0.1', '\alpha estimado', '\beta estimado');

toc

```

B.2. Código de la función para determinar la región de continuación (A_n, B_n) bajo horizonte H

Código para la obtención de la región de continuación (A_n, B_n). En la terminal de GNU Octave se debe invocar la función `RegionContinuacion(N, pH, t0, t1, Hi)`, donde $N \in \mathbb{N}$ es el máximo de observaciones por ser el caso truncado; $p_H \rightarrow 0$ es la tasa de fallos; t_0 es el parámetro bajo la hipótesis nula; t_1 es el parámetro bajo la hipótesis alternativa; Hi es la

indicación para generar las observaciones bajo la hipótesis nula si $H_i = 0$ o bien bajo la hipótesis alternativa si $H_i = 1$.

```
#Función que grafica la región de continuación en un experimento
#secuencial con observaciones de Bernoulli
#para probar una prueba de hipótesis simple y donde interviene un horizonte
#aleatorio H de distribución geométrica.
tic
function RegionDeContinuacion(N,pn,t0,t1,H_i)
# Cargamos el paquete statistics para hacer uso de la función binornd
pkg load statistics
format long
# Contantes debido a las observaciones de Bernoulli
k1 =(1-t1)/(1-t0); k2 = t1*(1-t0)/(t0*(1-t1));
# La condición  $L \leq c/(t1 - t0)$  se cumple con los parámetros
#propuestos y c , L abajo
c =0.0001; L =1;
# Observaciones bajo la hipótesis nula o alternativa
if H_i==0
    t = t0;
else
    t = t1;
end
# Con el siguiente bucle guardamos valores de la última etapa N y hacia atrás
for i = N:-1:1
    # Número de éxitos de 0 a i +1 , con i=N ,... ,1
    for j=1:(i+1)
        # Capturamos número de éxitos .
        Z(i,j) = k1^i*(k2^(j-1));
        # Registramos los valores para g de la regla de paro
        g =min(L,Z(i,j));
        # Registramos la función e la última etapa que coincide con g
        if i == N
            Rho(i,j)=g;
        else
            # Registramos la función rho en otras etapas distintas a N
            Rho(i,j) = g*pn+(1-pn)*min(g,c+Rho(i+1,j)*(1-t0)+Rho(i+1,j+1)*t0);
        end
    end
end
```

B.2. CÓDIGO DE LA FUNCIÓN PARA DETERMINAR LA REGIÓN DE CONTINUACIÓN (A_N, B_N)

```

    # Identificamos si nos encontramos en la región de ↵
    # continuación
    if g > c+Rho(i+1,j)*(1-t0)+Rho(i+1,j+1)*t0
        # Capturamos todos los valores que no paran el ↵
        # experimento
        Cont(i,j) = Z(i,j);
    end
end
end
# Identificamos los valores extremos mínimos del ↵
# experimento
Em(i) = Z(i,min(find(Z(i,:)>0)));
# Identificamos los valores extremos máximos del ↵
# experimento
EM(i) = max(Z(i,:));
end
# Evitar errores cuando no hay valores .
Cont(Cont==0) = NaN;
for i =1:length(Cont)
    # Identificamos región para no rechazar H0
    m(i) = Cont(i,min(find(Cont(i,:)>0)));
    # Identificamos región para rechazar H0
    M(i) = max(Cont(i,:));
end
# Número máximo de réplicas del método
# Montecarlo
MC =100;
x =1:1:N;
y =1:1:length(Cont) ;
for j =1:MC
    s=1;
    # Generación de observaciones aleatorias de distribución ↵
    # Bernoulli bajo H0 y
    # registro del tiempo de paro
    for i =1:length(Cont)
        s = k2^binornd(1,t)*k1*s;
        way(i,j) = s;
        if s <min(Cont(i,:))
            d(j) =0;
            break
        elseif s >max(Cont(i,:))
            d(j) =1;
            break
        else d(j) = NaN;
        endif
    end
end

```

```

    end
    finales(j) = i;
end
k = 1:length(way(:,MC/2));
disp(['Tiempo de paro promedio', num2str(mean(finales)) , '←
    con máximo de observaciones N= ', num2str(N)])
if H_i == 0
    disp(['Observaciones bajo H_0 con decisión correcta', ←
        num2str((length(d)-sum(d))/length(d)*100), '%'])
else
    disp(['Observaciones bajo H_1 con decisión correcta', ←
        num2str(sum(d)/length(d)*100), '%'])
end
# Mostramos los resultados gráficos
hold on
plot(x, log(Em(x)), 'r-', 'LineWidth', 3)
##plot(x, log(EM(x)), 'r - ', 'LineWidth', 3)
##plot(y, log(m(y)), 'b', 'LineWidth', 7)
##plot(y, log(M(y)), 'b', 'LineWidth', 7)
##plot(k, log(way), 'g', 'LineWidth', 2)
xlabel(('Tasa de fallos'), 'FontName', 'Arial', 'FontSize', ←
    28);
ylabel(('Promedio muestral bajo H_0 y H_1'), 'FontSize', 28);

endfunction

```

Bibliografía

- Akongo'o L., Christen, A. (2008). Bayesian sequential analysis for multiple-arm clinical trials, *Statistics and Computing* 19, 99–109. <https://link.springer.com/article/10.1007/s11222-008-9074-y>
- Bartky, W. (1943). Multiple Sampling with Constant Probability, *The Annals of Mathematical Statistics* 14: 363–377.
- Bartroff, J., Lai, T.Z., Shih, M. (2013). *Sequential Experimentation in Clinical Trials. Design and Analysis* USA. Springer.
- Bhattacharjee, D. and Mukhopadhyay, N. (2012). On SPRT and RSPRT for the Unknown Mean in Normal Distribution with Equal Mean and Variance. *Sequential Analysis: Design Methods and Applications* 31:1 108-134. DOI: 10.1080/07474946.2012.652015
- Burkholder, D. L. and Wijsman, R. A. (1963). Optimum Properties and Admissibility of Sequential Tests, *Annals of Mathematical Statistics* 34: 1–17.
- Chow, Y. S., Robbins, H., y Siegmund, D. (1971). *Great Expectations: the Theory of Optimal Stopping*. Boston: Houghton Mifflin Company.
- Closas, P., Coma, E., and Méndez, L. (2012). *Sequential detection of influenza epidemics by the Kolmogorov-Smirnov test*. BMC Medical Informatics & Decision Making. 12:112
- Cochlar, J. and Vrana, I. (1978). On the Optimum Sequential Test of two Hypotheses for Statistically Dependent Observations, *Kibernetika*, 14: 57–69.
- Cruz-Suárez, H., Ilhuicatzí-Roldán, R., and Montes-de-Oca, R. (2013). Markov Decision Processes on Borel Spaces With Total Cost and Random Horizon, *Journal of Optimization Theory and Applications* 162: 329–346.
- Dayanik, S. and Yu, A. J. (2013). Reward-Rate Maximization in Sequential Identification Under a Stochastic Deadline, *SIAM Journal on Control and Optimization* 51: 2922–2948.
- DeGroot, M. H. (1970). *Optimal Statistical Decisions*. USA. McGraw Hill, Inc.
- Dodge, H. F., and Romig, H. G. (1929). A Method of Sampling Inspection, *The Bell System Technical Journal* 8: 613–631.
- Eisenberg, B., Ghosh, B.K., and Simons, G. (1976). Properties of general sequential probability ratio tests, *Annals of Statistics* 4: 237–251.

- Fauß, M. & Zoubir, A. M. (2015). A Linear Programming Approach to Sequential Hypothesis Testing, *Sequential Analysis* 34:2, 235-263, DOI: 10.1080/07474946.2015.1030981.
- Ferguson, T. S. (1967). *Mathematical Statistics: a Decision Theoretic Approach*. New York: Academic Press.
- Frazier, P. I. and Yu, A. J. (2008). Sequential Hypothesis Testing under Stochastic Deadlines, *Advances in Neural Information Processing Systems*, 20. MIT Press, Cambridge, MA, 465–472.
- Iida, T. and Mori, M. (1996). Markov Decision Processes with Random Horizon. *Journal of the Operations Research Society of Japan*. 39, No. 4, 592-603.
- Karlin, S. and Taylor H.M. (1975). *A First Course in Stochastic Process*. New York; Wiley. Vol. 2. Academic Press. New York.
- Kiefer, J. and Weiss, L. (1957). Some Properties of Generalized Sequential Probability Ratio Tests, *Annals of Mathematical Statistics* 28: 57–75.
- Liu, Y. and Blostein, D. (1992). Optimality of The Sequential Probability Ratio Test for Nonstationary Observations, *IEEE Transactions on Information Theory* 28: 177–182.
- Lorden, G. (1980). Structure of Sequential Tests Minimizing an Expected Sample Size. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete (Probability Theory and Related Fields)*, 51: 291–302.
- Novikov, A. (2008). Optimal Sequential Tests for Two Simple Hypotheses Based on Independent Observations, *International Journal of Pure and Applied Mathematics* 45: 291–314.
- Novikov, A. (2009). Optimal Sequential Tests for Two Simple Hypotheses, *Sequential Analysis* 28: 188–217.
- Novikov A. & Palacios-Soto J.L. (2020) Sequential hypothesis tests under random horizon, *Sequential Analysis*, 39:2, 133-166, DOI: 10.1080/07474946.2020.176687.
- Puterman, M.L. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York: John Wiley & Sons Inc..
- Rockafellar, R. T. (1970). *Convex Analysis*. New Jersey: Princeton University Press.
- Samuel-Cahn, E. (1996). Optimal Stopping with Random Horizon with Application to the Full-Information Best-Choice Problem with Random Freeze, *Journal of the American Statistical Association* 91: 357-364.
- Schmitz, N. (1995). Comment on “Optimality of the sequential probability ratio test for nonstationary observations by Y. Liu and S.D. Blostein, *IEEE Transactions on Information Theory* 91: 861.
- Shiryayev, A. N. (1978). *Optimal Stopping Rules*. New York: Springer-Verlag.

- Wald, A. (1947). *Sequential Analysis*. New York; Wiley.
- Wald, A. (1950). *Statistical Decision Functions*. Canada. John Wiley & Sons, Inc.
- Wald, A. and Wolfowitz, J. (1948). Optimum Character of the Sequential Probability Ratio Test, *Annals of Mathematics Statistics* 19: 326–339.
- Weiss, L. (1962). Testing one simple hypothesis against another, *The Annals of Mathematical Statistics* 52: 273–281.
- Weiss, L. (1962). On Sequential Tests which Minimize the Maximum Expected Sample Size, *Journal of American Statistical Association* 57: 551–566.
- Zhang, W., Moustakides, G.V., and Poor, H.V. (2016). Opportunistic Detection Rules: Finite and Asymptotic Analysis, *IEEE Transaction on Information Theory* 62: 2140–2152.