

Universidad Autónoma Metropolitana - Iztapalapa  
División de Ciencias Básicas e Ingeniería

---

---



**OPTIMALIDAD DE PRUEBAS SECUENCIALES  
PARA EXPERIMENTOS CON  
HORIZONTE ALEATORIO**

TESIS QUE PRESENTA

**JUAN LUIS PALACIOS SOTO**

PARA OBTENER EL GRADO DE  
**MAESTRO EN CIENCIAS EN MATEMÁTICAS  
APLICADAS E INDUSTRIALES**

ASESOR: **DR. ANDREY NOVIKOV**

JURADO CALIFICADOR:

PRESIDENTE: **DR. JUAN GONZÁLEZ HERNÁNDEZ**

SECRETARIO: **DR. RAÚL MONTES DE OCA MACHORRO**

VOCAL: **DR. ANDREY NOVIKOV**

MÉXICO, D.F. MARZO 2014



# Resumen

En este trabajo desarrollamos la teoría para caracterizar la estructura de las pruebas secuenciales óptimas para experimentos, introduciendo un horizonte aleatorio, a diferencia de un horizonte infinito como en la teoría clásica. Analizamos las pruebas secuenciales óptimas encontradas con horizonte aleatorio en términos de razón de probabilidades y al observar un caso particular tenemos una relación con las clásicas *pruebas secuenciales de razón de probabilidades* (SPRT).

**Abstract:** In this work we developed the theory to characterize the structure of optimal sequential tests for experiments, introducing a random horizon, unlike an infinity horizon as in the classical theory. We analyzed the optimal sequential tests found with random horizon in terms of probability ratio and when a particular case is observed we have a relationship with the classical *sequential probability ratio tests* (SPRT).



*Este trabajo está dedicado principalmente a mi madre y a mi hermano  
quienes se adelantaron en este camino; así como a toda mi familia  
quienes me apoyan, aun estando lejos de ellos.  
Así mismo, a Liz A. Carrasco, que de no haber sido por ella...¡Este  
trabajo lo habría terminado hace más de medio año!*



# Agradecimientos

Primeramente agradezco a Dios, por darme la dicha de continuar en este mundo y permitirme conocer un  $\epsilon$  más de la matemática. Extiendo mi gratitud a mi asesor el Dr. Andrey Novikov, por su valiosa aportación y ayuda para el desarrollo de este trabajo. También deseo agradecer a mis sinodales, Raúl Montes de Oca y Juan González, por sus observaciones y correcciones, haciendo que este trabajo sea mejor. Así mismo, a mis amigos y compañeros de la maestría, en especial a: Héctor Manuel Garduño, por sus observaciones y ayuda en algunos pasajes de este trabajo y a Liz Arlette Carrasco, por su apoyo en momentos difíciles. Agradezco a mi padre por todo lo que ha sido en mi vida. Un agradecimiento muy especial a la Universidad Autónoma Metropolitana Unidad Iztapalapa, por haberme cobijado en estos estudios de posgrado.



# ÍNDICE GENERAL

|   |            |
|---|------------|
| <b>Resumen</b>  | <b>I</b>   |
| <b>Dedicatoria</b>  | <b>III</b> |
| <b>Agradecimientos</b>  | <b>V</b>   |
| <b>Introducción</b>   | <b>1</b>   |
| <b>1. Preliminares</b>  | <b>1</b>   |
| 1. Antecedentes . . . . .   | 2          |
| 2. Conceptos Básicos . . . . .  | 4          |
| 3. Pruebas de Hipótesis . . . . .   | 7          |
| 4. Planteamiento del Problema . . . . .   | 9          |
| <b>2. Optimalidad de Pruebas Secuenciales con Horizonte Aleatorio</b>               | <b>11</b>  |
| 1. Características y Desarrollo de las Pruebas Secuenciales con Horizonte Aleatorio | 11         |
| 2. Reducción al Problema de Paro Óptimo . . . . .                                   | 16         |
| 3. Paro Óptimo . . . . .  | 21         |
| 3.1. Caso Truncado . . . . .  | 22         |
| 3.2. Caso no Truncado . . . . .   | 32         |
| 4. Estructura de la Prueba Secuencial Óptima . . . . .                              | 43         |

|   |           |
|---|-----------|
| <b>3. Pruebas Óptimas y Razón de Probabilidades</b>                             | <b>45</b> |
| 1. Estructura de Pruebas Óptimas en Términos de Razón de Probabilidades . . . . | 45        |
| 2. Caso Particular: Función de Riesgo constante . . . . .                       | 61        |
| 3. Problema Inverso . . . . .   | 67        |
| 4. Conclusiones . . . . .   | 78        |
| <b>Bibliografía</b>   | <b>80</b> |

# Introducción

El análisis secuencial es un método de inferencia estadística que nació de los estudios realizados por Wald en 1947 [12] y se caracteriza porque el número de observaciones requeridas para terminar el experimento depende en cada paso de los resultados de las observaciones obtenidas previamente.

Una de las acciones del análisis secuencial es hacer pruebas de hipótesis estadísticas, donde el número de observaciones que se necesitan, en promedio, es sustancialmente menor e igualmente confiables que las pruebas tradicionales basadas en un número fijo de observaciones.

El estudio del análisis secuencial de Wald lo realizó suponiendo que el horizonte es infinito; es decir, no hay tiempo límite para terminar el experimento, pero el tiempo de paro es finito con probabilidad 1.

En el presente trabajo realizaremos la misma teoría que Wald, ahora introduciendo un horizonte aleatorio que puede ser acotado y que es independiente de los datos observados. Esto en la práctica tiene sentido toda vez que al estar realizando nuestro experimento puede llegar una orden o algún otro evento, independiente de los datos observados, que nos obligue a terminar de forma anticipada el experimento.

De las primeras aplicaciones del análisis secuencial, más en específico la prueba secuencial de razón de probabilidades y que veremos en el Capítulo 3, podemos mencionar que se ha utilizado en actos de espionaje [12]. Se ha aplicado y se sigue aplicando en procesos industriales, ensayos clínicos y actualmente también se encuentran aplicaciones a la neurociencia, por mencionar algunos [1, 3, 5].

El objetivo del Capítulo 1 es dar un panorama general de lo que es el análisis secuencial, revisando algunos antecedentes históricos, dando a conocer los conceptos básicos del análisis secuencial, basados principalmente en el trabajo de Novikov [9], como lo es la definición de prueba secuencial  $(\psi, \delta)$  con  $\psi$  conocida como regla de paro y  $\delta$  regla de decisión; así como el modelo matemático. Finalmente plantearemos el problema que queremos desarrollar en este trabajo.

En el Capítulo 2 planteamos la estructura de la prueba estadística secuencial que optimiza el número de observaciones con este enfoque del horizonte aleatorio. Trataremos en primer lugar las características y el desarrollo de las pruebas secuenciales con horizonte aleatorio. Enseguida veremos que el problema de la optimalidad de la prueba, dada la regla de decisión óptima, se reduce a encontrar la regla de paro óptima. Analizaremos la regla óptima en el caso truncado para después extenderla al caso general. Al final tendremos la estructura de la prueba secuencial óptima bajo el horizonte aleatorio.

Para el Capítulo 3, aplicaremos los resultados encontrados de las pruebas secuenciales óptimas, pero en términos de razón de probabilidades para ver qué estructura tiene bajo un horizonte aleatorio. Mostraremos una equivalencia entre las pruebas secuenciales óptimas del Capítulo 2 y una en términos de razón de probabilidades al encontrar cotas de continuación del experimento para cada etapa. Después veremos que al tomar como distribución del horizonte aleatorio la distribución geométrica tendremos, en cierto sentido, un método en particular desarrollado por Wald, la famosa *prueba secuencial de razón de probabilidades* (SPRT por sus siglas en inglés) [12] y en la que existen dos cotas de continuación del experimento. Finalmente, estudiaremos el problema inverso, el cual consiste en dar cotas de continuación y encontrar los parámetros iniciales que hacen de la prueba en cuestión  $(\psi, \delta)$ , la óptima.

# CAPÍTULO 1

## PRELIMINARES

En cualquier experimento estadístico es necesario recopilar datos para con ellos hacer inferencia de con qué distribución podrían venir éstos. Uno de los modelos más frecuentes de un experimento estadístico es una *muestra de una distribución de probabilidades*. Esto significa que se observan  $n$  datos de manera independiente y todos vienen con la misma distribución. Cuando se quiere una muestra de tamaño  $n$ , ésta se hace en una sola toma. La idea del análisis secuencial es ir formando la muestra de manera sucesiva en función de las mismas observaciones y terminando en el momento conveniente, de hecho en la práctica lo que nos interesa es terminar lo antes posible, lo cual se logra haciendo menos observaciones y que se traduce en menos recursos, tiempo, etc. Pero puede ocurrir algo ajeno a nuestra decisión que nos obligue a terminar de observar los datos. Por ejemplo, que venga una orden superior que nos exija terminar el experimento, ya sea por falta de recursos, de tiempo o por cualquier otro evento ajeno a nuestra decisión. A este suceso se le conoce como horizonte aleatorio acotado. También podríamos estar ante un horizonte no acotado, pero finito con probabilidad 1 o un horizonte infinito, como en la teoría clásica.

Para minimizar el número de observaciones definiremos una regla que nos indique en qué momento parar el experimento. En términos matemáticos podemos definir una función indicadora que nos señale tal situación.

El objetivo de este trabajo es crear una teoría alterna del análisis secuencial de la optimalidad de pruebas secuenciales, pero ahora introduciendo el concepto del horizonte aleatorio.

En este capítulo daremos el formalismo necesario al concepto del experimento secuencial,

restringido a las distribuciones discretas con un número finito de valores, suponiendo que las observaciones son independientes e idénticamente distribuidas (i.i.d.).

# 1

## Antecedentes

En muchos casos al hacer pruebas de hipótesis el número de observaciones; es decir, el número del tamaño de la muestra en la cual se basa la prueba, se trata regularmente de manera constante para muchos problemas en particular. El análisis secuencial tiene una característica esencial que se distingue de las clásicas pruebas de hipótesis y que consiste en que el número de observaciones que se requieren en la prueba secuencial depende de los resultados de las observaciones y por lo tanto es visto como una variable aleatoria. Además, el número de observaciones que se requieren para tomar una decisión sobre alguna de las pruebas de hipótesis es con mucha frecuencia dos o tres veces inferior a las muestras fijas de tamaño  $n$  e igual de confiables, de ahí su importancia porque se ahorra tiempo, esfuerzo, dinero, etc.

Wald fue el primero quien trabajó en el análisis estadístico secuencial. Definió dos reglas  $(\psi, \delta)$ , una regla de paro  $\psi$  (que nos dice el tamaño de la muestra) y una vez que hemos terminado el experimento tomar una decisión sobre una de las dos pruebas de hipótesis simples por medio de otra regla que llamó regla de decisión,  $\delta$ . Se han encontrado trabajos como el de Dayanik y Yu donde la decisión final se toma sobre  $m > 2$  pruebas de hipótesis [3].

Wald en su trabajo supuso que el fin del experimento llegaría sólo a través de la regla de paro y no por algún otro suceso o imprevisto, lo cual se traduce a que el horizonte sobre el que se trabajó es infinito. En este trabajo analizó el caso truncado, que se puede pensar que hay una fecha límite para tomar una decisión en el experimento secuencial; así como el caso no truncado y continuando con el horizonte infinito en ambos casos. Para encontrar la forma que debería tener la prueba secuencial estadística óptima, Wald trabajó sobre todas las pruebas que cumplieran con ciertas restricciones, las cuales tienen que ver con los errores conocidos en las pruebas de hipótesis como del tipo I y del tipo II que se cometen al rechazar la hipótesis nula cuando es cierta y aceptar la hipótesis nula cuando es falsa, respectivamente. Transformó el problema de minimización con restricciones a uno sin restricciones por medio de un enfoque Bayesiano.

Así encontró la estructura que deberían tener la regla de paro  $\psi$  y la regla de decisión  $\delta$  en la prueba óptima. Una vez que se obtiene la forma de la regla de decisión óptima, el problema se convierte en un problema de paro óptimo para minimizar la función objetivo.

Teniendo la forma de la prueba secuencial óptima  $(\psi^*, \delta^*)$ , Wald definió una nueva función con base en las funciones de probabilidades, bajo las hipótesis nula  $f_{\theta_0}(X_1, X_2, \dots, X_n)$  y la alternativa  $f_{\theta_1}(X_1, X_2, \dots, X_n)$ , que llamó *razón de probabilidades*, la cual calculó para ambos casos; es decir, tomando a  $f_{\theta_0}/f_{\theta_1}$  y su recíproco. Encontró dos cotas  $A$  y  $B$  ( $0 < A < B < \infty$ ), las cuales simplifican la prueba secuencial; es decir, modificó las reglas de paro y de decisión con base en la razón de probabilidades. El resultado más famoso del análisis secuencial es la optimalidad de las pruebas secuenciales en razón de probabilidades (SPRT), cuya demostración se presentó por primera vez en 1948 [13]. La demostración de la optimalidad de la SPRT es parecida a una prueba de hipótesis la cual está basada, a diferencia de la SPRT, en una muestra de tamaño fijo conocida como el Lema de Neyman-Pearson [8].

En el trabajo de Wald no consideró el costo unitario por observación. En cambio Dayanik y Yu en [3] introducen un costo, el cual se aplica por tomar una decisión tardía; es decir, como una multa por tomar la decisión antes que llegue el horizonte aleatorio. En este mismo trabajo se plantea un horizonte aleatorio donde se trata de encontrar la probabilidad desconocida,  $f(\cdot)$  de los datos, pero antes de que llegue el tiempo de paro aleatorio, el cual es desconocido, observable y que se supone con una distribución geométrica. Además, se asocia un costo con la regla admisible  $(\tau, \mu)$  (prueba secuencial).

El problema de optimalidad en caso truncado también lo abordan Bhattacharjee y Mukhopadhyay [1]. Además, exploran el tema de estimar intervalos de confianza para el parámetro  $\theta$  de los datos observados. Suponen que la densidad  $f_\theta$  pertenece a la familia exponencial. También se considera un costo conocido, asociado al tiempo de paro.

En el análisis secuencial clásico se suponen los datos independientes e idénticamente distribuidos (i.i.d), en cambio en el trabajo de Eisenberg, Ghosh y Simons [4], los datos no son necesariamente i.i.d.

Podemos encontrar trabajos como el de Frazier y Yu [5], en el que se propone un horizonte estocástico con distribución gamma. Para resolver la optimalidad de la prueba emplean herramientas de programación dinámica. También introducen un costo unitario  $c$  que asocian a cada unidad de tiempo, pero el costo es por la demora de tomar la decisión. Además, plantean un costo  $d$  al exceder el plazo límite; es decir, por tomar una decisión antes del paro. El objetivo en este trabajo es encontrar el costo mínimo cuando el plazo o fecha límite no ha llegado.

En los trabajos de Cruz Suárez, Ilhuicatzí y Montes de Oca, así como en el de Iida y Mori [2, 6], utilizan un horizonte aleatorio el cual llaman horizonte aleatorio de planeación el cual emplean en Procesos de Decisión de Markov (MDP por sus siglas en inglés). Bajo ciertos supuestos, muestran la existencia de la estrategia óptima para la MDP con horizonte aleatorio y la obtienen proponiendo un algoritmo que se resuelve numéricamente. Los resultados dependen de la distribución propuesta para el horizonte. Además en [2] se trata con costos que pueden ser no acotados.

Lorden en [7] resuelve el problema de optimización llevando el problema de minimizar la función objetivo con las restricciones de los errores tipo I y tipo II a un problema de minimización sin restricciones, incorporando las restricciones a la función objetivo, parecido al método de Lagrange, introduciendo tantos multiplicadores de Lagrange como restricciones se tengan, en este caso dos. No se supone un horizonte aleatorio y la probabilidad de que el paro sea finito es uno.

En el trabajo de Samuel-Cahn [11] se introduce un horizonte aleatorio y se aplica al llamado problema de la secretaria, el cual consiste en encontrar la mejor opción para ocupar una plaza. Suponen que los candidatos poseen ciertas cualidades y que éstas tienen una distribución conocida y continua. Tratan de encontrar al mejor elemento.

En las siguientes dos secciones veremos el modelo matemático, los conceptos básicos y repasaremos las pruebas de hipótesis en el contexto secuencial. En la última sección plantearemos el problema que queremos abordar en el presente trabajo.

## 2

### Conceptos Básicos

Empezaremos recordando algunos conceptos básicos que emplearemos para el desarrollo de este trabajo.

Sea  $X \in A$  una variable aleatoria que toma solamente un número finito de valores, con sus probabilidades respectivas  $f(x) = P(X = x)$  y  $A$  un conjunto finito. Al ser  $f$  una distribución de probabilidades, entonces

$$\sum_{x \in A} f(x) = 1.$$

Los datos experimentales  $X_1, X_2, X_3, \dots$  que serán observados de forma secuencial, son copias independientes de  $X$ , es decir, son observaciones independientes e idénticamente distribuidas (i.i.d.). Entonces,

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \prod_{i=1}^n f(x_i), \quad (1.1)$$

para todo  $n = 1, 2, 3, \dots$ . Por simplicidad denotaremos la parte derecha de (1.1) como  $f^n$ .

Necesitamos definir una regla para las observaciones que se obtienen, con el fin de indicar en qué momento el experimento debe parar. Para esto último, sea  $\psi_n = \psi_n(X_1, X_2, \dots, X_n)$  una función que toma sólo valores 0 ó 1. Esta será nuestra **función indicadora de paro**. Cualquier función de este tipo nos puede servir como la regla de paro en la etapa  $n$  siempre que la interpretemos de la siguiente manera:

- parar si  $\psi_n \equiv 1$ ,
- seguir observando si  $\psi_n \equiv 0$ .

Con esto podemos definir a

$$\psi = (\psi_1, \psi_2, \psi_3, \dots),$$

como **regla de paro**. Con esta regla tenemos el siguiente algoritmo del experimento secuencial:

1. Se toma la primera observación  $X_1$  y ponemos  $n = 1$ .
2. Si  $\psi_n(X_1, X_2, \dots, X_n) = 1$ , entonces terminamos el experimento,
3. en otro caso, observamos  $X_{n+1}$ , aumentamos  $n$  en 1 y regresamos al paso 2.

Entonces el experimento secuencial terminará cuando ocurra el primer momento  $n$  tal que  $\psi_n \equiv 1$ . De esta manera el número final de datos observados es

$$\nu_\psi = \begin{cases} \min\{1 \leq n \in \mathbb{N} : \psi_n(X_1, X_2, \dots, X_n) = 1\} & \text{si existe} \\ \infty & \text{en otro caso} \end{cases} \quad (1.2)$$

Pueden existir las reglas de paro para las cuales es posible que  $\nu_\psi$  sea infinito, por ejemplo, tomemos  $\psi_n \equiv 0 \forall n \geq 1$ , entonces  $\nu_\psi = \infty$ . Pero por cuestiones prácticas prescindiremos de ellas.

Al número aleatorio  $\nu_\psi$ , definido en (1.2) se le denomina *tiempo de paro*. La distribución de  $\nu_\psi$  está dada por

$$P(\nu_\psi = n) = P(\psi_1 = 0, \psi_2 = 0, \dots, \psi_{n-1} = 0, \psi_n = 1).$$

Podemos calcular esta probabilidad en términos de la regla de paro y de la distribución de  $X$ . Para hacer esto último, definamos la siguiente función indicadora

$$s_n^\psi = s_n^\psi(X_1, X_2, \dots, X_n) = (1 - \psi_1)(1 - \psi_2) \cdots (1 - \psi_{n-1})\psi_n, \quad (1.3)$$

para  $n = 1, 2, \dots$ . Observemos que  $s_n^\psi = 1$  si y sólo si  $\nu_\psi = n$  (ahora tiene sentido el subíndice  $\psi$  ya que  $\nu$  sólo depende de la regla de paro). Entonces,

$$\begin{aligned} P(\nu_\psi = n) &= E s_n^\psi(X_1, X_2, \dots, X_n) = E s_n^\psi \\ &= \sum_{x_1 \cdots x_n} s_n^\psi f^n(x_1, x_2, \dots, x_n) = \sum_{x_1 \cdots x_n} s_n^\psi f^n, \end{aligned}$$

esto último por (1.1).

Podemos hablar de  $E[\nu_\psi] = \sum_{n=1}^{\infty} nP(\nu_\psi = n)$ , siempre que  $P(\nu_\psi < \infty) = 1$  (como en [7]), porque de lo contrario la función de distribución tendría un salto al 1 debido al valor infinito que tomaría  $\nu_\psi$ . Entonces,

$$E[\nu_\psi] = \sum_{n=1}^{\infty} nP(\nu_\psi = n) = \sum_{n=1}^{\infty} n \sum_{x_1 \cdots x_n} s_n^\psi f^n.$$

Hasta ahora hemos caracterizado la forma del valor esperado del número final de datos observados y que depende de la regla de paro, sin considerar un horizonte aleatorio. En la siguiente sección veremos cómo entra en juego la regla de decisión para aceptar o rechazar alguna de las hipótesis simples en la prueba secuencial.

# 3

## Pruebas de Hipótesis

En la estadística inferencial hay dos grandes áreas de interés: la estimación de parámetros y las pruebas de hipótesis. En investigaciones experimentales, en algunas ocasiones, el objetivo principal es estimar parámetros siempre que haya incertidumbre de con qué distribución vengan los datos. Por ejemplo, si en el modelo clásico (1.1) la probabilidad  $f(x)$  depende de un parámetro  $\theta$ ,  $f(x) = f_\theta(x)$ , entonces cualquier pregunta relacionada con el valor verdadero de  $\theta$ , cuando se tienen sólo los datos de la muestra  $X_1, X_2, \dots, X_n$ , es una pregunta estadística.

Recordemos algunas definiciones básicas.

**Definición 1.1** (Hipótesis Estadística). *Una hipótesis estadística es una afirmación o conjetura acerca de la distribución de una o más variables aleatorias. Si la hipótesis estadística especifica completamente la distribución, entonces es llamada simple; de otra forma se llama compuesta.*

La definición anterior, nos dice que si los valores de los parámetros a estimar son únicos, entonces la hipótesis estadística se llama simple, de lo contrario se llamará compuesta. Pero nosotros trabajaremos con pruebas de hipótesis simples, por lo que tenemos la siguiente.

**Definición 1.2** (Prueba de una Hipótesis Estadística). *La prueba de una hipótesis estadística  $H$  es una regla o procedimiento para decidir si rechazamos a  $H$  o no.*

Por ejemplo, podemos suponer que tenemos una muestra de una distribución de Bernoulli, esto es

$$f(x) = P(X = x) = \begin{cases} p & \text{si } x = 1 \\ 1 - p & \text{si } x = 0. \end{cases}$$

El parámetro  $\theta$  es la probabilidad de “éxito”  $p \in [0, 1]$ .

En nuestro caso de la estadística secuencial, veremos el caso básico de las pruebas, como es el caso de dos hipótesis simples. Entonces, el problema que estudiaremos será, con base en los datos observados, decidir cuál de las dos hipótesis,  $H_0 : \theta = \theta_0$  o  $H_1 : \theta = \theta_1$ , es verdadera. Por

ejemplo, en el caso de la distribución de Bernoulli, una pregunta sería, ¿es cierto que los “éxitos” y “fracasos”, en la sucesión observada, vienen con la misma frecuencia? Esto es,  $H_0 : p = 0.5$ , y si existiera otra suposición sobre el valor de  $p$ , digamos  $H_1 : p = 0.6$ , tendremos que pensar en cómo los datos muestrales  $X_1, X_2, \dots, X_n$ , nos ayudarán a decidir a favor de una u otra de las hipótesis planteadas. Pero esto es para una muestra tomada de golpe.

Sabemos que al tratar con dos hipótesis simples, nula  $H_0$  y alternativa  $H_1$ , podemos cometer ciertos errores al tomar una mala decisión sobre alguna de las hipótesis. Entonces queremos que los errores que cometemos sean lo más pequeños.

**Definición 1.3.** *El rechazo de  $H_0$  cuando ésta es cierta, se llama error del tipo I y la aceptación de  $H_0$  cuando es falsa se llama error del tipo II. El tamaño del error del tipo I se define como la probabilidad de que éste ocurra y análogamente el tamaño del error del tipo II es la probabilidad de que este error ocurra.*

Así tenemos que el tamaño del error del tipo I es  $P_\theta(\text{Rechazar } H_0 | H_0 \text{ es cierta})$ , mientras que el tamaño del error del tipo II es  $P_\theta(\text{Aceptar } H_0 | H_0 \text{ es falsa})$ .

Ahora lo que nos interesa es implementar las pruebas de hipótesis en el contexto del experimento secuencial. Para esto necesitamos formalizar el concepto de un procedimiento de una prueba de hipótesis en este marco secuencial.

Como ya hemos mencionado anteriormente, nuestro experimento puede terminar en cualquier momento al tiempo  $n \geq 1$  y cuando eso suceda tenemos que decidir cuál de las dos hipótesis es con la que nos quedamos, cualquiera que sea  $n$ . Es decir, vamos a tomar una decisión de aceptar la  $H_0$  o rechazarla a favor de  $H_1$  después de que el experimento se detuvo. Entonces, como en el caso de la regla de paro, ahora necesitamos una regla de decisión para la prueba de hipótesis. Tomando en cuenta lo anterior, definamos una regla de decisión  $\delta$  como una sucesión de funciones indicadoras  $\delta_n = \delta_n(X_1, X_2, \dots, X_n)$ ,  $n = 1, 2, 3, \dots$ ,

$$\delta = (\delta_1, \delta_2, \delta_3, \dots, \delta_n, \dots),$$

con la siguiente interpretación al terminar el experimento secuencial en la etapa  $\nu_\psi = n$ :

- se acepta la hipótesis  $H_0$  si  $\delta_n = 0$  y
- se rechaza  $H_0$  a favor de  $H_1$  si  $\delta_n = 1$ .

Con lo anterior, tenemos la siguiente.

**Definición 1.4.** Una prueba secuencial estadística se define como una pareja  $(\psi, \delta)$ , donde  $\psi$  es una regla de paro y  $\delta$  es una regla de decisión.

Con la anterior definición podemos observar ahora que las pruebas clásicas, basadas en una muestra aleatoria de tamaño fijo  $n$ , es un caso particular de esta definición con  $\psi_1 \equiv \psi_2 \equiv \dots \equiv \psi_{n-1} \equiv 0$  y  $\psi_n \equiv 1$  y únicamente pueden ser identificadas con la regla de decisión  $\delta_n$ . Ya con los datos obtenidos  $X_1, X_2, \dots, X_n$  se realiza el experimento para aceptar la hipótesis  $H_i$ , con  $i = \delta_n$ .

Esta clase de pruebas es más variada, ya que cada prueba tiene además de la regla de decisión una regla de paro la cual determina el número final de observaciones que se realizaron en el experimento.

Ya que tenemos la definición de lo que se conoce como prueba estadística secuencial, seguiremos con el problema a resolver en el presente trabajo.

## 4

### Planteamiento del Problema

Como mencionamos anteriormente, Wald trabajó el análisis secuencial estadístico con un horizonte infinito. En este trabajo desarrollaremos una teoría alterna a lo hecho por Wald en [12], pero introduciendo un horizonte aleatorio, que puede hacer que nuestro experimento termine antes de que la regla de paro lo detenga. Esto tiene sentido práctico ya que al estar realizando nuestro experimento secuencial, puede suceder un evento de fuerza mayor que nos obligue a terminar de forma anticipada, ya sea porque se agotaron los recursos, vino una orden superior exigiendo su terminación, o cualquier otro suceso que ocurre de manera aleatoria.

Definiremos a nuestro horizonte como una variable aleatoria que denotaremos con la letra  $D$ . Entonces, uno de los objetivos principales de este trabajo es caracterizar la estructura de la prueba secuencial óptima  $(\psi, \delta)$  con la incorporación de este horizonte  $D$  y ver si, como en el caso clásico, minimizamos el número final de datos observados en la prueba bajo la hipótesis nula, teniendo como restricciones los errores del tipo I ( $\alpha$ ) y tipo II ( $\beta$ ) que se cometen al realizar pruebas de hipótesis. Además, vamos a introducir un costo unitario  $c \geq 0$  al ir observando los datos bajo la hipótesis nula. Así, el objetivo es minimizar el costo total, el cual está dado por

$$cE_{\theta_0}[\nu_\psi \wedge D],$$

sobre todas las pruebas secuenciales  $(\psi, \delta)$  restringido a  $\alpha(\psi, \delta) \leq \alpha$  y  $\beta(\psi, \delta) \leq \beta$ .

Para tal propósito utilizaremos, en esencia, el método de multiplicadores de Lagrange como Lorden en [7], para poder incorporar las restricciones de los errores en la función objetivo a minimizar. Por lo tanto, el problema de minimización con restricciones se convierte en un problema de minimización sin restricciones, es decir,

$$\min_{\psi, \delta} \left\{ cE_{\theta_0}[\nu_\psi \wedge D] + \lambda_0 \alpha(\psi, \delta) + \lambda_1 \beta(\psi, \delta) \right\}.$$

Para llegar a la regla óptima analizaremos como Wald el caso truncado, que es cuando la regla tiene un máximo de observaciones  $N$ . Por lo que la regla de paro que se trata en este caso es simplemente

$$\psi = (\psi_1, \psi_2, \psi_3, \dots, \psi_{N-1}, \psi_N \equiv 1).$$

Después abordaremos, con ayuda del caso truncado, el caso no truncado.

Pondremos los resultados de la prueba óptima en el contexto de razón de probabilidades con el objeto de encontrar cotas  $A$  y  $B$  (en caso de que existan bajo el horizonte aleatorio) de la región de continuación, a partir de constantes  $c, \lambda_0, \lambda_1 \geq 0$ .

Finalmente, trataremos el caso inverso en caso de horizonte aleatorio; es decir, a partir de las cotas  $A$  y  $B$ , hallar tres constantes  $c, \lambda_0, \lambda_1 \geq 0$  que hacen la prueba óptima.

## CAPÍTULO 2

# OPTIMALIDAD DE PRUEBAS SECUENCIALES CON HORIZONTE ALEATORIO

## 1

### Características y Desarrollo de las Pruebas Secuenciales con Horizonte Aleatorio

En términos muy generales, lo que queremos realizar en esta parte es introducir el horizonte aleatorio en los errores del tipo I y II en pruebas de hipótesis simples, ya que al realizar una prueba, se genera un algoritmo de ir observando los datos y finalmente se llega a la conclusión una vez terminado el experimento. Sin embargo, la conclusión puede ser errónea y ésta viene dada por los errores antes mencionados. Esto nos servirá para después desarrollar el problema de paro óptimo con horizonte aleatorio  $D$  cuando realizamos pruebas de hipótesis sobre los datos observados al tiempo  $n$  ( $x_1, x_2, \dots, x_n$ ). Por el momento no pondremos condición sobre el tipo de distribución que pueda tener el horizonte aleatorio  $D$ , que bien puede ser del tipo geométrica o uniforme como en [11].

Veamos qué ocurre cuando hacemos pruebas de hipótesis y cómo entra en este contexto el horizonte aleatorio  $D$ .

Para tener una calidad en la inferencia estadística, lo conveniente es procurar que los errores

no sean frecuentes. Para medir esta frecuencia se usa la probabilidad. Por lo tanto, tenemos que tomar en cuenta las probabilidades de los errores bajo el horizonte aleatorio  $D$ . Entonces y por lo que vimos en el capítulo anterior, sabemos que la probabilidad del error tipo I es  $\alpha(\psi, \delta) = P_{\theta_0}(\text{rechazar } H_0)$  y la probabilidad del error tipo II es  $\beta(\psi, \delta) = P_{\theta_1}(\text{aceptar } H_0)$ .

Es conveniente definir las siguientes funciones indicadoras.

$$t_n^\psi = (1 - \psi_1)(1 - \psi_2) \cdots (1 - \psi_{n-1}) \quad (2.1)$$

y  $t_1^\psi \equiv 1$  para toda regla de paro  $\psi$ . Observemos que  $t_n^\psi$  es una sucesión decreciente de números en  $[0,1]$ ,

$$\begin{aligned} t_1^\psi &= 1, \\ t_2^\psi &= (1 - \psi_1) \leq 1 = t_1^\psi \quad \text{porque cada } \psi_i \text{ es } 0 \text{ ó } 1, \\ t_3^\psi &= (1 - \psi_2)t_2^\psi \leq t_2^\psi, \\ &\vdots \\ t_{n+1}^\psi &= (1 - \psi_n)t_n^\psi \leq t_n^\psi, \\ &\vdots \end{aligned}$$

además,

$$\begin{aligned} t_n^\psi - t_{n+1}^\psi &= (1 - \psi_1)(1 - \psi_2) \cdots (1 - \psi_n) - (1 - \psi_1)(1 - \psi_2) \cdots (1 - \psi_{n-1}), \\ &= (1 - \psi_1)(1 - \psi_2) \cdots (1 - \psi_{n-1}(1 - (1 - \psi_n))), \\ &= (1 - \psi_1)(1 - \psi_2) \cdots (1 - \psi_{n-1})\psi_n, \\ &= s_n^\psi, \end{aligned}$$

con  $s_n^\psi$  dada en (1.3). Notemos que  $s_n^\psi = 1 \Leftrightarrow \psi_i = 0 \ \forall \ i = 1, \dots, n-1$  y  $\psi_n = 1$ , pero esto ocurre  $\Leftrightarrow t_i^\psi = 1$  para toda  $i = 1, \dots, n$ . Además, observemos que  $t_n^\psi \psi_n = s_n^\psi$ . Así, para  $D$  como horizonte aleatorio y  $\nu_\psi$  como en (1.2), tenemos que

$$\begin{aligned} P_{\theta_0}(\text{rechazar } H_0) &= P_{\theta_0}(\text{rechazar } H_0, D > \nu_\psi) + P_{\theta_0}(\text{rechazar } H_0, D \leq \nu_\psi) \\ &= \sum_{n=1}^{\infty} P_{\theta_0}(\delta_n = 1, s_n^\psi = 1, D > n) \\ &\quad + \sum_{d=1}^{\infty} P_{\theta_0}(\delta_d = 1, \{\psi_1 = 0, \psi_1 = 0, \dots, \psi_{d-1} = 0\}, D = d) \end{aligned}$$

$$\begin{aligned}
 &= \sum_{n=1}^{\infty} E_{\theta_0}[\delta_n s_n^\psi] P(D > n) + \sum_{d=1}^{\infty} P_{\theta_0}(\delta_d = 1, t_d = 1) P(D = d) \\
 &= \sum_{n=1}^{\infty} E_{\theta_0}[\delta_n s_n^\psi] P(D > n) + \sum_{d=1}^{\infty} E_{\theta_0}[\delta_d t_d^\psi] P(D = d) \\
 &= \sum_{n=1}^{\infty} \sum_{x_1 \cdots x_n} s_n^\psi \delta_n f_{\theta_0}^n P(D > n) \\
 &\quad + \sum_{d=1}^{\infty} \sum_{x_1 \cdots x_d} t_d^\psi \delta_d f_{\theta_0}^d P(D = d).
 \end{aligned}$$

Por lo tanto,

$$\alpha(\psi, \delta) = \sum_{n=1}^{\infty} \sum_{x_1 \cdots x_n} \delta_n f_{\theta_0}^n \left( s_n^\psi P(D > n) + t_n^\psi P(D = n) \right). \quad (2.2)$$

Ahora veamos qué sucede para el caso de aceptar  $H_0$ .

$$\begin{aligned}
 P_{\theta_1}(\text{aceptar } H_0) &= P_{\theta_1}(\text{aceptar } H_0, D > \nu_\psi) + P_{\theta_1}(\text{aceptar } H_0, D \leq \nu_\psi) \\
 &= \sum_{n=1}^{\infty} P_{\theta_1}(\delta_n = 0, s_n^\psi = 1, D > n) \\
 &\quad + \sum_{d=1}^{\infty} P_{\theta_1}(\delta_d = 0, \{\psi_1 = 0, \psi_1 = 0, \dots, \psi_{d-1} = 0\}, D = d) \\
 &= \sum_{n=1}^{\infty} E_{\theta_1}[(1 - \delta_n) s_n^\psi] P(D > n) + \sum_{d=1}^{\infty} E_{\theta_1}[(1 - \delta_d) t_d^\psi] P(D = d) \\
 &= \sum_{n=1}^{\infty} \sum_{x_1 \cdots x_n} s_n^\psi (1 - \delta_n) f_{\theta_1}^n P(D > n) \\
 &\quad + \sum_{d=1}^{\infty} \sum_{x_1 \cdots x_d} t_d^\psi (1 - \delta_d) f_{\theta_1}^d P(D = d).
 \end{aligned}$$

Luego,

$$\beta(\psi, \delta) = \sum_{n=1}^{\infty} \sum_{x_1 \cdots x_n} (1 - \delta_n) f_{\theta_1}^n \left( s_n^\psi P(D > n) + t_n^\psi P(D = n) \right). \quad (2.3)$$

Sea  $\nu_\psi \wedge D = \min\{\nu_\psi, D\}$ . Como  $\nu_\psi$  es una variable aleatoria, entonces su valor esperado puede ser calculado tanto bajo la hipótesis  $H_0$  como bajo la  $H_1$  y al introducir el horizonte aleatorio  $D$ , tenemos dos características:

$$N(\theta_0; \psi) := E_{\theta_0}[\nu_\psi \wedge D] \quad \text{y} \quad N(\theta_1; \psi) := E_{\theta_1}[\nu_\psi \wedge D],$$

las cuales tienen que ver con la regla de paro y con el horizonte  $D$  mas no con la regla de decisión. Podemos tener el caso en que el experimento no termine y esto puede ocurrir si el horizonte aleatorio no es acotado y  $\nu_\psi = \infty$ . Por cuestiones prácticas y por el momento, pediremos que el experimento termine en algún momento, ya sea porque  $\nu_\psi < \infty$  y/o  $D$  sea acotado, en términos probabilísticos lo podemos ver como  $P(\nu_\psi \wedge D < \infty) = 1$ .

Supongamos que el parámetro verdadero de  $\nu_\psi \wedge D$  es  $\theta_0$ , entonces nos podemos enfocar al cálculo de  $N(\theta_0; \psi)$ . Así,

$$\begin{aligned} N(\theta_0; \psi) = E_{\theta_0}(\nu_\psi \wedge D) &= E_{\theta_0}(\nu_\psi \wedge D)I_{\{D > \nu_\psi\}} + E_{\theta_0}(\nu_\psi \wedge D)I_{\{D \leq \nu_\psi\}} \\ &= E_{\theta_0}(\nu_\psi)I_{\{D > \nu_\psi\}} + E_{\theta_0}(D)I_{\{D \leq \nu_\psi\}} \\ &= \sum_{n=1}^{\infty} nP_{\theta_0}(\nu_\psi = n, D > n) + \sum_{n=1}^{\infty} nP_{\theta_0}(D = n, n \leq \nu_\psi) \\ &= \sum_{n=1}^{\infty} nP_{\theta_0}(\nu_\psi = n)P(D > n) + \sum_{n=1}^{\infty} nP(D = n)P_{\theta_0}(n \leq \nu_\psi) \\ &= \sum_{n=1}^{\infty} nE_{\theta_0}(s_n^\psi)P(D > n) + \sum_{n=1}^{\infty} nP(D = n)E_{\theta_0}(t_n^\psi) \\ &= \sum_{n=1}^{\infty} \sum_{x_1 \cdots x_n} n s_n^\psi f_{\theta_0}^n P(D > n) \\ &\quad + \sum_{n=1}^{\infty} nP(D = n) \sum_{x_1 \cdots x_n} t_n^\psi f_{\theta_0}^n \\ &= \sum_{n=1}^{\infty} \sum_{x_1 \cdots x_n} n f_{\theta_0}^n t_n^\psi \left( \psi_n P(D > n) + P(D = n) \right). \end{aligned} \tag{2.4}$$

Hasta ahora hemos caracterizado el problema. Lo que ahora sigue es introducir el costo por dato observado

Al realizar un experimento estadístico secuencial representa un costo para quien o quienes lo realizan, por lo que es conveniente ver el problema en estos términos. Así, sea  $c \geq 0$  el costo unitario de las observaciones, entonces el costo total promedio del experimento secuencial depende de la regla de paro  $\psi$  y del horizonte aleatorio  $D$ , el cual está dado por

$$C(\psi; c) := E_{\theta_0}[c(\nu_\psi \wedge D)] = cN(\theta_0; \psi). \quad (2.5)$$

La anterior función la llamaremos función de costo.

El problema del análisis secuencial es hallar la forma de la prueba que minimiza el número promedio de observaciones en la clase de todas las pruebas cuyas probabilidades de error no excedan ciertas cotas. Sea  $\Delta(\alpha, \beta)$  la clase de las pruebas  $(\psi, \delta)$  tales que satisfacen las siguientes desigualdades:

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

siendo  $\alpha$  y  $\beta$  dos números reales,  $0 < \alpha, \beta < 1$ , generalmente muy pequeños. Entonces el objetivo es hallar una prueba en  $\Delta(\alpha, \beta)$  con  $C(\psi; c)$  mínimo en esta clase.

El problema planteado representa un problema de optimización con restricciones siendo  $C(\psi; c)$  la función objetivo. Para solución de problemas de optimización con restricciones, como ya mencionamos anteriormente, se conoce el método de multiplicadores de Lagrange [7]. En esencia, es el método de Lagrange el que emplearemos para minimizar la función objetivo, pero en el espacio de funciones. Este método consiste en reducir el problema de optimización con restricciones a uno sin restricciones, incorporando las restricciones en una nueva función objetivo, conocida como función de Lagrange. Entonces la función que nos interesa tiene la forma:

$$L(\psi, \delta; c, \lambda_0, \lambda_1) = C(\psi; c) + \lambda_0\alpha(\psi, \delta) + \lambda_1\beta(\psi, \delta),$$

donde  $\lambda_0, \lambda_1 \geq 0$  son conocidas como multiplicadores de Lagrange.

Sustituyendo  $C(\psi; c)$  como en (2.5), tenemos finalmente que la función de Lagrange tiene la forma:

$$L(\psi, \delta; c, \lambda_0, \lambda_1) = cE_{\theta_0}[\nu_\psi \wedge D] + \lambda_0\alpha(\psi, \delta) + \lambda_1\beta(\psi, \delta). \quad (2.6)$$

Esta forma de la función de Lagrange es la que trabajaremos en la mayoría de las ocasiones para demostrar ciertos teoremas, pero cabe hacer mención que no será la única, como veremos más adelante.

Nos interesa encontrar una solución (si la hay) no trivial con al menos un dato observado ( $E_{\theta_0}[\nu_\psi \wedge D] \geq 1$ ). El caso trivial es precisamente cuando  $\lambda_0 = 0$  o  $\lambda_1 = 0$ . Esto es cierto toda vez que si  $\lambda_0 = 0$  en (2.6), tenemos que  $P_{\theta_0}(\text{rechazar } H_0)$  no contribuye en la función de Lagrange, por lo que podemos rechazar siempre a la hipótesis nula, es decir,  $P_{\theta_0}(\text{rechazar } H_0) = 1$ . Entonces

se cumplirá que  $P_{\theta_1}(\text{aceptar } H_0) = 0$ . Se sigue que la función de Lagrange a minimizar con  $\lambda_0 = 0$  es  $L(\psi, \delta; c, 0, \lambda_1) = cE_{\theta_0}[\nu_\psi \wedge D]$ . Por lo tanto, dadas  $\psi, \delta$  se cumple

$$\min_{\psi, \delta} L(\psi, \delta; c, 0, \lambda_1) = c.$$

De igual manera se cumple para  $\lambda_1 = 0$ . Entonces para no tener el caso trivial vamos a suponer que  $\lambda_0, \lambda_1 > 0$ .

Como  $c, \lambda_0$  y  $\lambda_1$  permanecen constantes y por cuestiones prácticas abusaremos de la notación, haciendo  $L(\psi, \delta) = L(\psi, \delta; c, \lambda_0, \lambda_1)$  y  $C(\psi) = C(\psi; c)$ .

## 2

### Reducción al Problema de Paro Óptimo

El siguiente teorema representa la reducción del problema con restricciones a uno sin ellas aplicando el método de Lagrange. Por lo tanto haremos uso de la función de Lagrange que vimos al final de la sección anterior.

**Teorema 2.1.** *Sea  $(\psi^*, \delta^*)$  una prueba tal que para cualquier otra prueba secuencial  $(\psi, \delta)$*

$$L(\psi^*, \delta^*) \leq L(\psi, \delta), \tag{2.7}$$

y

$$\alpha(\psi^*, \delta^*) = \alpha \quad \beta(\psi^*, \delta^*) = \beta. \tag{2.8}$$

*Entonces para cualquier prueba  $(\psi, \delta)$  que satisfice*

$$\alpha(\psi, \delta) \leq \alpha \quad \beta(\psi, \delta) \leq \beta \tag{2.9}$$

*se cumple*

$$C(\psi^*) \leq C(\psi). \tag{2.10}$$

*La desigualdad (2.10) es estricta si alguna de las desigualdades de (2.9) es estricta.*

Observemos que las condiciones (2.8) garantizan que  $(\psi^*, \delta^*) \in \Delta(\alpha, \beta)$ , así que (2.10) representa la optimalidad de  $(\psi^*, \delta^*)$  en la clase  $\Delta(\alpha, \beta)$  desde el punto de vista promedio del costo.

*Demostración.*

Sea  $(\psi^*, \delta^*)$  que satisface (2.7) y (2.8). Supongamos que  $(\psi, \delta)$  es una prueba arbitraria que satisface (2.9), entonces,

$$\begin{aligned} C(\psi^*) + \lambda_0\alpha + \lambda_1\beta &= C(\psi^*) + \lambda_0\alpha(\psi^*, \delta^*) + \lambda_1\beta(\psi^*, \delta^*), \\ &= L(\psi^*, \delta^*), \\ &\leq L(\psi, \delta), \end{aligned} \tag{2.11}$$

$$\begin{aligned} &= C(\psi) + \lambda_0\alpha(\psi, \delta) + \lambda_1\beta(\psi, \delta), \\ &\leq C(\psi) + \lambda_0\alpha + \lambda_1\beta. \end{aligned} \tag{2.12}$$

Luego,

$$C(\psi^*) \leq C(\psi).$$

Lo que demuestra la primera parte del teorema. Además, es claro que si una de las desigualdades en (2.9) es estricta, entonces la desigualdad en (2.12) es estricta. Ahora si  $C(\psi^*) = C(\psi)$  para alguna prueba  $(\psi, \delta)$ , entonces las desigualdades (2.11) y (2.12) son en realidad igualdades. Por lo tanto,  $\alpha(\psi, \delta) = \alpha$  y  $\beta(\psi, \delta) = \beta$ , y esto concluye la demostración. ■

Del Teorema 2.1 vemos que para hallar la forma de la prueba óptima necesitamos minimizar la función de Lagrange  $L(\psi, \delta)$ . Para esto, daremos una regla de decisión  $\delta^*$  óptima en el sentido de que para cualquier otra regla de decisión  $\delta$ , se cumple

$$L(\psi, \delta^*) \leq L(\psi, \delta).$$

Para dar a  $\delta^*$  necesitamos introducir la siguiente notación. Sea

$$\delta_n^* = I_{\{\lambda_0 f_0^n - \lambda_1 f_1^n \leq 0\}} = \begin{cases} 1 & \text{si } \lambda_0 f_0^n - \lambda_1 f_1^n \leq 0, \\ 0 & \text{en otro caso.} \end{cases} \tag{2.13}$$

Entonces hagamos  $\delta^* = (\delta_1^*, \delta_2^*, \dots, \delta_n^*, \dots)$ .

**Teorema 2.2.** *Sea  $\delta^* = (\delta_1^*, \delta_2^*, \dots)$  una regla de decisión con  $\delta_i^*$  como en (2.13). Entonces para cualquier regla de paro  $\psi$  y cualquier regla de decisión  $\delta$*

$$L(\psi, \delta^*) \leq L(\psi, \delta).$$

Además, para el valor de  $L(\psi, \delta^*)$  se tiene la siguiente expresión

$$L(\psi, \delta^*) = C(\psi) + \sum_{n=1}^{\infty} \sum t_n^\psi (\psi_n P(D > n) + P(D = n)) \min\{\lambda_0 f_0^n, \lambda_1 f_1^n\},$$

con  $t_n^\psi$  dado como en (2.1).

*Demostración.*

Observemos que por la definición de la función de Lagrange dada en (2.6), sólo hay que probar que

$$\lambda_0 \alpha(\psi, \delta^*) + \lambda_1 \beta(\psi, \delta^*) \leq \lambda_0 \alpha(\psi, \delta) + \lambda_1 \beta(\psi, \delta) \quad (2.14)$$

y por consiguiente

$$\lambda_0 \alpha(\psi, \delta^*) + \lambda_1 \beta(\psi, \delta^*) = \sum_{n=1}^{\infty} \sum t_n^\psi (\psi_n P(D > n) + P(D = n)) \min\{\lambda_0 f_0^n, \lambda_1 f_1^n\}.$$

Para probar esto, haremos uso del siguiente.

**Lema 2.3.** *Sea  $x$  una variable con un número finito de valores y sea  $F(x)$  cualquier función de  $x$ . Entonces para cualesquiera funciones  $\phi(x), s(x), 0 \leq \phi(x), s(x) \leq 1$ , se tiene*

$$\sum s(x) F(x) I_{\{F(x) \leq 0\}} \leq \sum s(x) \phi(x) F(x). \quad (2.15)$$

*Demostración.*

Observemos que (2.15) es equivalente a probar que

$$0 \leq \sum s(x) F(x) [\phi(x) - I_{\{F(x) \leq 0\}}]. \quad (2.16)$$

Veamos que  $F(x) [\phi(x) - I_{\{F(x) \leq 0\}}] \geq 0$  para todo  $x$ .

- Si  $F(x) \leq 0$  entonces  $\phi(x) - I_{\{F(x) \leq 0\}} = \phi(x) - 1 \leq 0$ , por lo tanto su producto es no negativo.

- Si  $F(x) > 0$  entonces  $\phi(x) - I_{\{F(x) \leq 0\}} = \phi(x) \geq 0$ , y nuevamente su producto es no negativo.

De esta manera vemos que todos los sumandos de (2.16) son en efecto no negativos, con lo que se demuestra la desigualdad (2.15) ■

Regresando a la prueba de (2.14), representamos el lado derecho utilizando (2.2) y (2.3). Entonces

$$\begin{aligned} \lambda_0\alpha(\psi, \delta) + \lambda_1\beta(\psi, \delta) &= \sum_{n=1}^{\infty} \sum_{x_1 \cdots x_n} \lambda_0 \delta_n f_0^n \left( s_n^\psi P(D > n) + t_n^\psi P(D = n) \right) \\ &\quad + \sum_{n=1}^{\infty} \sum_{x_1 \cdots x_n} \lambda_1 (1 - \delta_n) f_1^n \left( s_n^\psi P(D > n) + t_n^\psi P(D = n) \right) \\ &= \sum_{n=1}^{\infty} \sum_{x_1 \cdots x_n} \left( \lambda_0 \delta_n f_0^n + \lambda_1 (1 - \delta_n) f_1^n \right) t_n^\psi \left( \psi_n P(D > n) + P(D = n) \right). \end{aligned}$$

Pongamos por el momento a  $H_\psi^n = \psi_n P(D > n) + P(D = n)$ , entonces

$$\begin{aligned} \lambda_0\alpha(\psi, \delta) + \lambda_1\beta(\psi, \delta) &= \sum_{n=1}^{\infty} \sum_{x_1 \cdots x_n} (\lambda_0 \delta_n f_0^n + \lambda_1 (1 - \delta_n) f_1^n) t_n^\psi H_\psi^n \\ &= \sum_{n=1}^{\infty} \sum_{x_1 \cdots x_n} (\lambda_0 \delta_n f_0^n - \lambda_1 \delta_n f_1^n + \lambda_1 f_1^n) t_n^\psi H_\psi^n \\ &= \sum_{n=1}^{\infty} \sum_{x_1 \cdots x_n} (\lambda_0 f_0^n - \lambda_1 f_1^n) \delta_n t_n^\psi H_\psi^n + \lambda_1 \sum_{n=1}^{\infty} \sum_{x_1 \cdots x_n} f_1^n t_n^\psi H_\psi^n. \end{aligned}$$

Observemos que a los primeros sumandos de este último resultado podemos aplicar el Lema 2.3, con  $\phi = \delta_n$ ,  $s = t_n^\psi H_\psi^n$  y  $F = \lambda_0 f_0^n - \lambda_1 f_1^n$ , entonces tenemos que

$$\lambda_0\alpha(\psi, \delta) + \lambda_1\beta(\psi, \delta) \geq \sum_{n=1}^{\infty} \sum_{x_1 \cdots x_n} (\lambda_0 f_0^n - \lambda_1 f_1^n) t_n^\psi H_\psi^n I_{\{\lambda_0 f_0^n - \lambda_1 f_1^n \leq 0\}} \quad (2.17)$$

$$+ \lambda_1 \sum_{n=1}^{\infty} \sum_{x_1 \cdots x_n} f_1^n t_n^\psi H_\psi^n \quad (2.18)$$

$$\begin{aligned}
 &= \sum_{n=1}^{\infty} \sum_{x_1 \cdots x_n} (\lambda_0 f_0^n - \lambda_1 f_1^n) t_n^\psi H_\psi^n \delta_n^* + \lambda_1 \sum_{n=1}^{\infty} \sum_{x_1 \cdots x_n} f_1^n t_n^\psi H_\psi^n \\
 &= \lambda_0 \alpha(\psi, \delta^*) + \lambda_1 \beta(\psi, \delta^*)
 \end{aligned}$$

Por lo tanto,  $L(\psi, \delta) \geq L(\psi, \delta^*)$  como se quería probar.

Notemos que de (2.17) y (2.18) tenemos

$$\begin{aligned}
 \lambda_0 \alpha(\psi, \delta^*) + \lambda_1 \beta(\psi, \delta^*) &= \sum_{n=1}^{\infty} \sum_{x_1 \cdots x_n} \left[ (\lambda_0 f_0^n - \lambda_1 f_1^n) I_{\{\lambda_0 f_0^n - \lambda_1 f_1^n \leq 0\}} + \lambda_1 f_1^n \right] t_n^\psi H_\psi^n \\
 &= \sum_{n=1}^{\infty} \sum_{x_1 \cdots x_n} \min\{\lambda_0 f_0^n, \lambda_1 f_1^n\} t_n^\psi H_\psi^n.
 \end{aligned}$$

Lo anterior es cierto toda vez que

$$(\lambda_0 f_0^n - \lambda_1 f_1^n) I_{\{\lambda_0 f_0^n - \lambda_1 f_1^n \leq 0\}} + \lambda_1 f_1^n = \begin{cases} \lambda_0 f_0^n \leq \lambda_1 f_1^n & \text{si } I \equiv 1, \\ \lambda_1 f_1^n < \lambda_0 f_0^n & \text{si } I \equiv 0. \end{cases}$$

Por lo tanto, siempre tenemos el mínimo. Luego,

$$L(\psi, \delta^*) = C(\psi) + \sum_{n=1}^{\infty} \sum_{x_1 \cdots x_n} \min\{\lambda_0 f_0^n, \lambda_1 f_1^n\} t_n^\psi \left( \psi_n P(D > n) + P(D = n) \right). \quad (2.19)$$

■

Con el resultado de (2.19) queda claro que  $L(\psi, \delta^*)$  dependerá sólo de la regla de paro, así que, abusando nuevamente de la notación hagamos

$$L(\psi) = \inf_{\delta} L(\psi, \delta) = L(\psi, \delta^*).$$

Resulta de esta manera que el problema de minimización de  $L(\psi, \delta)$  se convierte en un problema de paro óptimo, ya que la función de Lagrange sólo depende de la regla de paro. Por lo tanto, el problema se centra en encontrar la regla de paro  $\psi^*$  para la cual se cumple

$$L(\psi^*) \leq L(\psi)$$

para cualquier regla de paro  $\psi$  con las restricciones de los errores del tipo I y tipo II. Lo anterior es cierto, ya que por el Teorema 2.2 y para cualquier prueba  $(\psi, \delta)$ , se cumple que

$$L(\psi, \delta) \geq L(\psi, \delta^*) = L(\psi) \geq L(\psi^*) = L(\psi^*, \delta^*).$$

En consecuencia,  $L(\psi^*, \delta^*)$  será la prueba óptima.

Con esta nueva notación, la función de Lagrange a minimizar la podemos ver como

$$L(\psi) = cE_{\theta_0}[\nu_\psi \wedge D] + \lambda_0\alpha(\psi, \delta^*) + \lambda_1\beta(\psi, \delta^*), \quad (2.20)$$

o bien por el mismo Teorema 2.2

$$L(\psi) = cE_{\theta_0}[\nu_\psi \wedge D] + \sum_{n=1}^{\infty} \sum_{x_1 \cdots x_n} \min\{\lambda_0 f_0^n, \lambda_1 f_1^n\} t_n^\psi \left( \psi_n P(D > n) + P(D = n) \right). \quad (2.21)$$

Además, si denotamos  $m_n = \min\{\lambda_0 f_0^n, \lambda_1 f_1^n\}$  y sustituyendo a  $E_{\theta_0}[\nu_\psi \wedge D]$  en (2.21) por (2.4), tenemos finalmente una forma más explícita.

$$\begin{aligned} L(\psi) &= c \sum_{n=1}^{\infty} \sum_{x_1 \cdots x_n} n f_0^n t_n^\psi \left( \psi_n P(D > n) + P(D = n) \right) \\ &\quad + \sum_{n=1}^{\infty} \sum_{x_1 \cdots x_n} m_n t_n^\psi \left( \psi_n P(D > n) + P(D = n) \right) \\ &= \sum_{n=1}^{\infty} \sum_{x_1 \cdots x_n} t_n^\psi \left( \psi_n P(D > n) + P(D = n) \right) \left( cn f_0^n + m_n \right). \end{aligned} \quad (2.22)$$

Esta última forma es la que usaremos en la mayor parte de lo que resta de este capítulo, pero habrá ocasiones que nos convendrá hacer uso de la forma (2.21).

No perdamos de vista que la función  $L(\psi)$  tiene, de manera implícita, los parámetros  $c, \lambda_0$  y  $\lambda_1$ , de igual manera  $m_n$  tiene a los parámetros  $\lambda_0, \lambda_1$  y los datos observados  $X_1, X_2, \dots, X_n$ .

### 3

## Paro Óptimo

En esta sección caracterizaremos la prueba secuencial óptima, primero para el caso truncado al tiempo  $N$  y después el problema de la optimalidad de la prueba lo extenderemos al caso

no truncado. En el caso truncado hablaremos de un caso más general de las reglas de paro a nivel truncado  $N$  como  $\psi = (\psi_1, \psi_2, \dots, \psi_{N-1}, \psi_N \equiv 1)$ , dejando que  $\psi_n$  sea cualquiera para  $n = 1, 2, \dots, N - 1$ .

### 3.1

#### Caso Truncado

En lo siguiente, resolveremos el problema de paro óptimo planteado anteriormente, cuando estamos en la clase de tiempos de paro truncado.

En algunas demostraciones haremos uso de la función de Lagrange de la forma  $L(\psi; c, \lambda_0, \lambda_1)$  con el fin de hacer más fácil de entender ciertas demostraciones, pero cuando sólo denotemos a  $L(\psi)$ , se entenderá que contiene a los parámetros  $c, \lambda_0$  y  $\lambda_1$ .

Sea  $N$  el número máximo de observaciones que se pueden realizar en la prueba; es decir,  $\psi_N \equiv 1$  y  $\psi_n$  arbitraria para toda  $1 \leq n \leq N - 1$ . Entonces de (2.22) podemos definir lo siguiente:

$$\begin{aligned}
 L_N(\psi) &= \sum_{n=1}^N \sum_{x_1 \cdots x_n} t_n^\psi \left( \psi_n P(D > n) + P(D = n) \right) \left( cnf_0^n + m_n \right) \\
 &= \sum_{n=1}^{N-1} \sum_{x_1 \cdots x_n} t_n^\psi \left( \psi_n P(D > n) + P(D = n) \right) \left( cnf_0^n + m_n \right) \\
 &\quad + \sum_{x_1 \cdots x_N} t_N^\psi \left( \psi_N P(D > N) + P(D = N) \right) \left( cNf_0^N + m_N \right) \\
 &= \sum_{n=1}^{N-1} \sum_{x_1 \cdots x_n} t_n^\psi \left( \psi_n P(D > n) + P(D = n) \right) \left( cnf_0^n + m_n \right) \\
 &\quad + \sum_{x_1 \cdots x_N} t_N^\psi P(D \geq N) \left( cNf_0^N + m_N \right),
 \end{aligned} \tag{2.23}$$

que es la función de Lagrange truncada al nivel  $N$  ( $\psi_N \equiv 1$ ).

Cabe hacer mención que cuando tengamos una suma de la siguiente forma:  $\sum_{i=1}^0 a_i = 0$ , su valor es cero, debido a que no hay elementos en la suma. Por ejemplo, si  $N = 1$  en la parte derecha de (2.23), entonces sólo quedará el segundo sumando.

Establezcamos el siguiente lema, el cual absorbe la mayor parte técnica para el paro óptimo en el caso truncado.

**Lema 2.4.** *Sea  $\psi = (\psi_1, \psi_2, \dots)$  cualquier regla de paro fija. Supongamos que  $r$  es un número entero,  $2 \leq r \leq N$ , y  $v_r = v_r(x_1, \dots, x_r)$  una función cualquiera. Entonces*

$$\begin{aligned}
 & \sum_{n=1}^{r-1} \sum_{x_1 \cdots x_n} t_n^\psi \left( \psi_n P(D > n) + P(D = n) \right) \left( cnf_0^n + m_n \right) \\
 & + \sum_{x_1 \cdots x_r} t_r^\psi P(D \geq r) \left( crf_0^r + v_r \right) \\
 \geq & \sum_{n=1}^{r-2} \sum_{x_1 \cdots x_n} t_n^\psi \left( \psi_n P(D > n) + P(D = n) \right) \left( cnf_0^n + m_n \right) \\
 & + \sum_{x_1 \cdots x_{r-1}} t_{r-1}^\psi P(D \geq r-1) \left( c(r-1)f_0^{r-1} + v_{r-1} \right),
 \end{aligned} \tag{2.24}$$

en donde

$$v_{r-1} = \min \left\{ m_{r-1}, p_{r-1} m_{r-1} + (1 - p_{r-1}) \left[ cf_0^{r-1} + \sum_{x_r} v_r(x_1, x_2, \dots, x_r) \right] \right\},$$

siendo  $p_{r-1} = \frac{P(D=r-1)}{P(D \geq r-1)}$ , siempre que  $P(D \geq r-1) > 0$ , de lo contrario en (2.24) tendremos sólo la suma hasta  $r-2$  en ambos lados.

La desigualdad en (2.24) se convierte en igualdad para

$$\psi_{r-1} = \begin{cases} I_{\{m_{r-1} \leq cf_0^{r-1} + \sum_{x_r} v_r\}} & \text{si } p_{r-1} < 1. \\ 1 & \text{si } p_{r-1} = 1. \end{cases}$$

*Demostración.*

Es suficiente probar que

$$\begin{aligned}
 & \sum_{x_1 \cdots x_{r-1}} t_{r-1}^\psi \left( \psi_{r-1} P(D > r-1) + P(D = r-1) \right) \left( c(r-1)f_0^{r-1} + m_{r-1} \right) \\
 & + \sum_{x_1 \cdots x_r} t_r^\psi P(D \geq r) \left( crf_0^r + v_r \right) \\
 \geq & \sum_{x_1 \cdots x_{r-1}} t_{r-1}^\psi P(D \geq r-1) \left( c(r-1)f_0^{r-1} + v_{r-1} \right).
 \end{aligned} \tag{2.25}$$

Recordando que  $t_r^\psi = t_{r-1}^\psi(1 - \psi_{r-1})$ , encontramos que la parte izquierda de (2.25) es igual a

$$\begin{aligned}
 & \sum_{x_1 \cdots x_{r-1}} t_{r-1}^\psi \left\{ \left( \psi_{r-1} P(D > r-1) + P(D = r-1) \right) \left( c(r-1)f_0^{r-1} + m_{r-1} \right) \right\} \\
 & + \sum_{x_1 \cdots x_{r-1}} t_{r-1}^\psi \sum_{x_r} (1 - \psi_{r-1}) P(D \geq r) (cf_0^r + v_r) \\
 = & \sum_{x_1 \cdots x_{r-1}} t_{r-1}^\psi \left\{ \psi_{r-1} P(D > r-1) (c(r-1)f_0^{r-1} + m_{r-1}) + P(D = r-1) (c(r-1)f_0^{r-1} + m_{r-1}) \right. \\
 & \left. + (1 - \psi_{r-1}) P(D \geq r) (cf_0^{r-1} + \sum_{x_r} v_r) \right\} \\
 = & \sum_{x_1 \cdots x_{r-1}} t_{r-1}^\psi \left\{ \psi_{r-1} P(D > r-1) cf_0^{r-1} - \psi_{r-1} P(D > r-1) cf_0^{r-1} + \psi_{r-1} P(D > r-1) m_{r-1} \right. \\
 & \left. + P(D = r-1) (c(r-1)f_0^{r-1} + m_{r-1}) + P(D \geq r) (cf_0^{r-1} + \sum_{x_r} v_r) \right. \\
 & \left. - \psi_{r-1} P(D \geq r) cf_0^{r-1} - \psi_{r-1} P(D \geq r) \sum_{x_r} v_r \right\} \\
 = & \sum_{x_1 \cdots x_{r-1}} t_{r-1}^\psi \left\{ \psi_{r-1} \left[ P(D > r-1) m_{r-1} - P(D > r-1) cf_0^{r-1} - P(D \geq r) \sum_{x_r} v_r \right] \right. \\
 & \left. + P(D = r-1) (c(r-1)f_0^{r-1} + m_{r-1}) + P(D \geq r) cf_0^{r-1} + P(D \geq r) \sum_{x_r} v_r \right\} \\
 = & \sum_{x_1 \cdots x_{r-1}} t_{r-1}^\psi \left\{ \psi_{r-1} \left[ P(D > r-1) (m_{r-1} - (cf_0^{r-1} + \sum_{x_r} v_r)) \right] + P(D \geq r) \sum_{x_r} v_r \right. \\
 & \left. + P(D = r-1) (c(r-1)f_0^{r-1} + m_{r-1}) + P(D \geq r) c(r-1)f_0^{r-1} + P(D \geq r) cf_0^{r-1} \right\} \\
 = & \sum_{x_1 \cdots x_{r-1}} t_{r-1}^\psi \left\{ \psi_{r-1} \left[ P(D > r-1) (m_{r-1} - (cf_0^{r-1} + \sum_{x_r} v_r)) \right] + P(D \geq r) (cf_0^{r-1} + \sum_{x_r} v_r) \right. \\
 & \left. + P(D \geq r-1) c(r-1)f_0^{r-1} + P(D = r-1) m_{r-1} \right\}. \tag{2.26}
 \end{aligned}$$

Aplicando el Lema 2.3 con

$$s = t_{r-1}^\psi,$$

$$\phi = \psi_{r-1} \text{ y}$$

$$F = P(D > r-1) (m_{r-1} - (cf_0^{r-1} + \sum_{x_r} v_r)),$$

tenemos que la parte derecha de (2.26) cumple que

$$\geq \sum_{x_1 \cdots x_{r-1}} t_{r-1}^\psi \left\{ \left[ P(D > r-1) (m_{r-1} - (cf_0^{r-1} + \sum_{x_r} v_r)) \right] \times \right.$$

$$\begin{aligned}
 & I_{\{P(D > r-1)(m_{r-1} - (cf_0^{r-1} + \sum_{x_r} v_r)) \leq 0\}} \\
 & + P(D \geq r)(cf_0^{r-1} + \sum_{x_r} v_r) + P(D \geq r-1)c(r-1)f_0^{r-1} + P(D = r-1)m_{r-1} \} \\
 = & \sum_{x_1 \cdots x_{r-1}} t_{r-1}^\psi \left\{ \min \left[ P(D > r-1)m_{r-1}, P(D > r-1)(cf_0^{r-1} + \sum_{x_r} v_r) \right] \right. \\
 & \left. + P(D \geq r-1)c(r-1)f_0^{r-1} + P(D = r-1)m_{r-1} \right\}. \tag{2.27}
 \end{aligned}$$

Supongamos que  $P(D > r-1) > 0$ , entonces el mínimo de (2.27) se puede simplificar como  $P(D > r-1) \min \{m_{r-1}, cf_0^{r-1} + \sum_{x_r} v_r\}$ . Además, (2.26) es igual a (2.27) si

$$\psi_{r-1} = I_{\{m_{r-1} \leq cf_0^{r-1} + \sum_{x_r} v_r\}}$$

Pero además, se cumple que

$$P(D = r-1) < P(D > r-1) + P(D = r-1) = P(D \geq r-1),$$

lo que implica

$$p_{r-1} = \frac{P(D = r-1)}{P(D \geq r-1)} < 1.$$

Por otra parte, si  $P(D > r-1) = 0$  y  $P(D = r-1) > 0$ , entonces

$$p_{r-1} = \frac{P(D = r-1)}{P(D \geq r-1)} = \frac{P(D = r-1)}{P(D > r-1) + P(D = r-1)} = \frac{P(D = r-1)}{P(D = r-1)} = 1$$

y (2.26) es igual a (2.27) si  $\psi_{r-1} \equiv 1$ .

En caso de que  $P(D \geq r-1) = 0$ , entonces  $P(D > k) = 0$  para toda  $k \geq r$ . Luego, (2.26) es igual a (2.27) y ambos son cero. Por lo tanto, tendríamos que verificar si la desigualdad se cumple para  $r-2$  y  $r-3$ .

Ahora, si el mínimo en (2.27) es  $P(D > r-1)m_{r-1}$ , la expresión entre llaves de (2.27) es igual a

$$\begin{aligned}
 & P(D > r-1)m_{r-1} + P(D \geq r-1)c(r-1)f_0^{r-1} + P(D = r-1)m_{r-1} \\
 = & P(D \geq r-1)(c(r-1)f_0^{r-1} + m_{r-1}) \\
 = & P(D \geq r-1)(c(r-1)f_0^{r-1} + v_{r-1}).
 \end{aligned}$$

En cambio, si el mínimo en (2.27) es  $P(D > r - 1)(cf_0^{r-1} + \sum_{x_r} v_r)$ , entonces la expresión respectiva en (2.27) es igual a

$$\begin{aligned}
 & P(D > r - 1)(cf_0^{r-1} + \sum_{x_r} v_r) + P(D = r - 1)m_{r-1} + P(D \geq r - 1)c(r - 1)f_0^{r-1} \\
 = & P(D \geq r - 1) \left\{ c(r - 1)f_0^{r-1} + \frac{P(D = r - 1)}{P(D \geq r - 1)} m_{r-1} \right. \\
 & \left. + \frac{P(D \geq r - 1) - P(D = r - 1)}{P(D \geq r - 1)} \left( cf_0^{r-1} + \sum_{x_r} v_r \right) \right\} \\
 = & P(D \geq r - 1) \left\{ c(r - 1)f_0^{r-1} + p_{r-1}m_{r-1} + (1 - p_{r-1}) \left( cf_0^{r-1} + \sum_{x_r} v_r \right) \right\} \\
 = & P(D \geq r - 1) \{ c(r - 1)f_0^{r-1} + v_{r-1} \}.
 \end{aligned}$$

Se sigue que (2.27) es igual a

$$\sum_{x_1 \cdots x_{r-1}} t_{r-1}^\psi P(D \geq r - 1) \left( c(r - 1)f_0^{r-1} + v_{r-1} \right).$$

Por lo tanto, hemos probado (2.25). ■

Observación I: En el anterior lema hemos trabajado con  $p_r = \frac{P(D=r)}{P(D \geq r)}$ , que se conoce como función de riesgo o también conocida como tasa de fallos. Notemos que si  $p_k = 1$ , entonces  $\psi_k \equiv 1$  y  $k$  es única con esta propiedad. Esto es cierto toda vez que si

$$p_k = \frac{P(D = k)}{P(D \geq k)} = \frac{P(D = k)}{P(D > k) + P(D = k)} = 1$$

entonces  $P(D > k) = 0$ . Pero esto ocurre si y sólo si  $P(D = n) = 0$  para toda  $n > k$ . Así podemos ver que por arriba de  $k$  no existe otro entero no negativo con esa propiedad. Supongamos que por abajo existe  $k_2 < k$ , tal que  $p_{k_2} = 1$ , entonces  $P(D > k_2) = 0$ . Luego, tendremos que  $P(D = k) = 0$  y con ello  $p_k$  queda ¡indefinido! Por lo tanto,  $k$  es único que hace  $p_k = 1$ .

Observación II: De las funciones  $v_n$  obtenemos la funciones indicadoras de  $\psi_n$  para la regla de paro, siempre que  $p_n < 1$ .

Ahora bien, el Lema 2.4 se puede aplicar inmediatamente a la función  $L_N(\psi)$  de (2.23), definiendo  $v_N = m_N$ . Entonces,

$$L_N(\psi) \geq \sum_{n=1}^{N-2} \sum_{x_1 \cdots x_n} t_n^\psi \left( \psi_n P(D > n) + P(D = n) \right) (cnf_0^n + m_n)$$

$$+ \sum_{x_1 \cdots x_{N-1}} t_{N-1}^\psi P(D \geq N-1) \left( c(N-1)f_0^{N-1} + v_{N-1} \right), \quad (2.28)$$

donde  $v_{N-1} = \min \left\{ m_{N-1}, p_{N-1}m_{N-1} + (1-p_{N-1})(cf_0^{N-1} + \sum_{x_N} v_N) \right\}$ , teniendo en cuenta que  $p_{N-1} = \frac{P(D=N-1)}{P(D \geq N-1)}$ . La igualdad se dará en (2.28) con

$$\psi_{N-1} = \begin{cases} I_{\{m_{N-1} \leq cf_0^{N-1} + \sum_{x_N} v_N\}} & \text{si } p_{N-1} < 1. \\ 1 & \text{si } p_{N-1} = 1. \end{cases} \quad (2.29)$$

Aplicando nuevamente el Lema 2.4 al lado derecho de (2.28), vemos que

$$\begin{aligned} L_N(\psi) &\geq \sum_{n=1}^{N-3} \sum_{x_1 \cdots x_n} t_n^\psi \left( \psi_n P(D > n) + P(D = n) \right) (cnf_0^n + m_n) \\ &\quad + \sum_{x_1 \cdots x_{N-2}} t_{N-2}^\psi P(D \geq N-2) \left( c(N-2)f_0^{N-2} + v_{N-2} \right), \end{aligned} \quad (2.30)$$

donde  $v_{N-2} = \min \left\{ m_{N-2}, p_{N-2}m_{N-2} + (1-p_{N-2})(cf_0^{N-2} + \sum_{x_{N-1}} v_{N-1}) \right\}$ . Teniendo la igualdad en (2.30) con

$$\psi_{N-2} = \begin{cases} I_{\{m_{N-2} \leq cf_0^{N-2} + \sum_{x_{N-1}} v_{N-1}\}} & \text{si } p_{N-2} < 1. \\ 1 & \text{si } p_{N-2} = 1, \end{cases} \quad (2.31)$$

siempre que  $p_{N-1} < 1$ , de lo contrario se elimina el caso cuando  $p_{N-2} = 1$ .

Aplicando de forma iterativa el Lema 2.4 hasta  $r = 2$ , vemos que la primera expresión de la parte derecha de 2.24 es nula, luego la cota será de la forma

$$L_N(\psi) \geq \sum_{x_1} (cf_0^1 + v_1), \quad (2.32)$$

además de algunas otras condiciones sobre la regla de paro  $\psi$  para tener igualdad en (2.28), (2.30) y (2.32). Entonces para tener igualdades en (2.28), (2.30) y (2.32) y si existe  $k < N$  tal que  $p_k = 1$ , entonces  $\psi_k \equiv 1$  y  $\psi_{k-1}, \psi_{k-2}, \dots, \psi_1$  análogos a la indicadora de (2.29); mientras que  $\psi_{k+1}, \psi_{k+2}, \dots, \psi_N$ , pueden ser cualquier función, en caso de que existan. En cambio, si  $p_k < 1$  para toda  $k < N$ , entonces  $\psi_k$  son indicadoras, comenzando con las indicadoras de (2.29)

y (2.31). Esto significa que estas condiciones describen la regla de paro óptima para el caso truncado a nivel  $N$  con horizonte aleatorio  $D$ .

El resultado formal se encuentra en el siguiente.

**Teorema 2.5.** *Sea  $\psi$  cualquier regla de paro truncada ( $\psi_N \equiv 1$ ). Entonces para cualquier  $1 \leq r \leq N - 1$  se cumplen las siguientes desigualdades*

$$\begin{aligned}
 L_N(\psi) &\geq \sum_{n=1}^r \sum_{x_1 \cdots x_n} t_n^\psi \left( \psi_n P(D > n) + P(D = n) \right) \left( cn f_0^n + m_n \right) \\
 &\quad + \sum_{x_1 \cdots x_{r+1}} t_{r+1}^\psi P(D \geq r + 1) \left( c(r + 1) f_0^{r+1} + V_{r+1}^N \right) \\
 &\geq \sum_{n=1}^{r-1} \sum_{x_1 \cdots x_n} t_n^\psi \left( \psi_n P(D > n) + P(D = n) \right) \left( cn f_0^n + m_n \right) \\
 &\quad + \sum_{x_1 \cdots x_r} t_r^\psi P(D \geq r) \left( cr f_0^r + V_r^N \right), \tag{2.33}
 \end{aligned}$$

en donde, por definición,  $V_N^N = m_N$ , y  $V_n^N$  son definidas para  $n = N - 1, N - 2, \dots, 1$  de la siguiente manera recursiva:

$$V_n^N = \min \left\{ m_n, p_n m_n + (1 - p_n) \left( cf_0^n + \sum_{x_{n+1}} V_{n+1}^N \right) \right\}, \tag{2.34}$$

con  $p_n = \frac{P(D=n)}{P(D \geq n)}$ .

Además, la cota inferior (2.33) se alcanza bajo las siguientes condiciones.

$$\psi_n = \begin{cases} I_{\{m_n \leq cf_0^n + \sum_{x_{n+1}} V_{n+1}^N\}} & \text{si } p_n < 1, \\ & n = 1, 2, \dots, N - 1, \\ 1 & \text{si } p_n = 1. \end{cases} \tag{2.35}$$

Si  $p_n = 1$ , comenzamos con  $V_n^N = m_n$  y  $V_k^N$  como en (2.34) para  $k = n - 1, \dots, r$ .

En el teorema anterior, hemos considerado que  $P(D \geq n) > 0$  para  $n = N - 1, \dots, 1$ . Ahora bien, si  $n = \min\{t : P(D \geq t) = 0\}$ , entonces tenemos que  $p_{n-1} = 1$ , toda vez que

$$p_{n-1} = \frac{P(D = n - 1)}{P(D \geq n - 1)} = \frac{P(D = n - 1)}{P(D \geq n) + P(D = n - 1)} = \frac{P(D = n - 1)}{P(D = n - 1)} = 1.$$

Por lo tanto, este caso queda contemplado dentro de (2.35).

Notemos que también las funciones  $V_n^N$  definidas en la ecuación (2.34) están relacionadas con la regla de paro dada en (2.35). Esto es cierto ya que si el mínimo es  $m_n$ , entonces se tiene la desigualdad:

$$m_n \leq m_n p_n + (1 - p_n)(c f_0^n + \sum_{x_{n+1}} V_{n+1}^N).$$

Esto es equivalente a tener

$$(1 - p_n)m_n \leq (1 - p_n)(c f_0^n + \sum_{x_{n+1}} V_{n+1}^N).$$

De lo anterior vemos que si  $p_n < 1$ , entonces tenemos la función indicadora de la función  $\psi_n$  de la regla de paro. En cambio si  $p_n = 1$ , entonces se sigue de (2.34) que  $V_n^N = m_n$ .

Ahora, hagamos  $r = 1$  en el Teorema 2.5, con lo que obtendremos la estructura de la regla de paro óptima en forma del siguiente corolario.

**Corolario 2.6.** *Sea  $N \geq 2$  un número entero y supongamos que  $p_n < 1$  para toda  $n = 1, 2, \dots, N-1$ .*

*Sean las funciones  $V_n^N = V_n^N(x_1, \dots, x_n)$ ,  $n = N-1, \dots, 1$ , definidas por las fórmulas recursivas (2.34) partiendo de  $V_N^N \equiv m_N$ .*

*Entonces para cualquier regla de paro  $\psi$  truncada a nivel  $N$  (es decir, tal que  $\psi_N \equiv 1$ ) se cumple la siguiente desigualdad:*

$$L_N(\psi) \geq c + \sum_x V_1^N(x). \quad (2.36)$$

*La cota (2.36) es alcanzada si  $\psi$  satisface (2.35) para toda  $n = N-1, \dots, 1$ , y  $\psi_N \equiv 1$ .*

*Por otra parte,  $\psi_n \equiv 1$  siempre que  $p_n = 1$  y partiendo de  $V_n^N \equiv m_n$ .*

Notemos que las funciones  $V_n^N$  definidas en (2.34) contienen, de manera implícita, los parámetros  $c, \lambda_0, \lambda_1$  y por supuesto al horizonte aleatorio  $D$  por medio de la función de riesgo  $p_r$ . Por lo que al hacer uso de estas funciones y para no ser redundantes, se entenderá que contiene a los parámetros antes mencionados.

El siguiente resultado es un caso particular cuando  $c = 0$  y que nos será de utilidad en el último capítulo tomando un horizonte  $D$  no acotado, aunque finito con probabilidad 1.

**Corolario 2.7.** Sea  $N \geq 2$  un número entero y supongamos que  $p_n < 1$  para toda  $n = 1, 2, \dots, N-1$ . Entonces dada una regla de paro  $\psi$  truncada a nivel  $N$  (es decir, tal que  $\psi_N \equiv 1$ ) se cumple la siguiente desigualdad:

$$L_N(\psi; 0, \lambda, \lambda_1) \geq \left( \lambda_0 P_{\theta_0}(\lambda_0 f_0^N \leq \lambda_1 f_1^N) + \lambda_1 P_{\theta_1}(\lambda_0 f_0^N > \lambda_1 f_1^N) \right) P(D > N) \\ + \sum_{n=1}^N \left( \lambda_0 P_{\theta_0}(\lambda_0 f_0^n \leq \lambda_1 f_1^n) + \lambda_1 P_{\theta_1}(\lambda_0 f_0^n > \lambda_1 f_1^n) \right) P(D = n). \quad (2.37)$$

La cota (2.37) es alcanzada si  $\psi = (\psi_1 \equiv 0, \psi_2 \equiv 0, \dots, \psi_{N-1} \equiv 0, \psi_N \equiv 1)$ . Además, las funciones  $V_n$  tienen la forma siguiente: para  $V_N^N = m_N$ , mientras para  $n = N-1, N-2, \dots, 1$  se tiene que

$$V_n^N = m_n p_n + (1 - p_n) \sum_{x_{n+1}} V_{n+1}^N. \quad (2.38)$$

*Demostración.*

Consideremos la función de Lagrange dada en (2.20), pero truncada a nivel  $N$ . Entonces para  $c = 0$ ,  $\delta^*$  como la regla de decisión óptima y tomando las expresiones para  $\alpha$  y  $\beta$  dadas en (2.2) y (2.3), respectivamente, tenemos que

$$L_N(\psi; 0, \lambda_0, \lambda_1) = 0E_{\theta_0}[\nu_\psi \wedge D] + \lambda_0 \alpha(\psi, \delta^*) + \lambda_1 \beta(\psi, \delta^*) \\ = \lambda_0 \sum_{n=1}^N \sum_{x_1 \dots x_n} \delta_n^* f_0^n (s_n^\psi P(D > n) + t_n^\psi P(D = n)) \\ + \lambda_1 \sum_{n=1}^N \sum_{x_1 \dots x_n} (1 - \delta_n^*) f_1^n (s_n^\psi P(D > n) + t_n^\psi P(D = n)).$$

Para esta regla de paro óptima  $\psi = (\psi_1 \equiv 0, \psi_2 \equiv 0, \dots, \psi_{N-1} \equiv 0, \psi_N \equiv 1)$ , se tiene que

$$t_n^\psi = (1 - \psi_1)(1 - \psi_2) \cdots (1 - \psi_{n-1}) = 1 \quad \text{para toda } n = 1, 2, \dots, N,$$

mientras que

$$s_n^\psi = (1 - \psi_1)(1 - \psi_2) \cdots (1 - \psi_{n-1})\psi_n = 0 \quad \text{para } n = 1, 2, \dots, N-1 \text{ y } s_N^\psi = 1.$$

Con lo anterior y dada cualquier otra regla de paro  $\psi$  truncada a nivel  $N$  se tiene

$$\begin{aligned}
L_N(\psi; 0, \lambda_0, \lambda_1) &\geq \lambda_0 \sum_{x_1 \cdots x_N} \delta_N^* f_0^N P(D > N) + \lambda_0 \sum_{n=1}^N \sum_{x_1 \cdots x_n} \delta_n^* f_0^n P(D = n) \\
&\quad + \lambda_1 \sum_{x_1 \cdots x_N} (1 - \delta_N^*) f_1^N P(D > N) + \lambda_1 \sum_{n=1}^N \sum_{x_1 \cdots x_n} (1 - \delta_n^*) f_1^n P(D = n) \\
&= \lambda_0 E_{\theta_0}[\delta_N^*] P(D > N) + \lambda_0 \sum_{n=1}^N E_{\theta_0}[\delta_n^*] P(D = n) \\
&\quad + \lambda_1 E_{\theta_1}[1 - \delta_N^*] P(D > N) + \lambda_1 \sum_{n=1}^N E_{\theta_1}[1 - \delta_n^*] P(D = n) \\
&= \lambda_0 P_{\theta_0}(\lambda_0 f_0^N \leq \lambda_1 f_1^N) P(D > N) + \lambda_0 \sum_{n=1}^N P_{\theta_0}(\lambda_0 f_0^n \leq \lambda_1 f_1^n) P(D = n) \\
&\quad + \lambda_1 P_{\theta_1}(\lambda_0 f_0^N > \lambda_1 f_1^N) P(D > N) + \lambda_1 \sum_{n=1}^N P_{\theta_1}(\lambda_0 f_0^n > \lambda_1 f_1^n) P(D = n).
\end{aligned}$$

Reagrupando la parte derecha de la igualdad anterior se cumple (2.37).

Para demostrar (2.38) y por la definición de  $V_n^N$  dada en (2.34) es suficiente verificar que  $V_n^N \leq m_n$  aplicando inducción inversa sobre  $n = N - 1, N - 2, \dots, 1$ , pero primero notemos que se cumple  $\sum_{x_{n+1}} m_{n+1} \leq m_n$  para toda  $n = N - 1, N - 2, \dots, 1$  o equivalentemente

$$\sum_{x_{n+1}} \min\{\lambda_0 f_0^{n+1}, \lambda_1 f_1^{n+1}\} \leq \min\{\lambda_0 f_0^n, \lambda_1 f_1^n\}. \quad (2.39)$$

Lo cual es inmediato toda vez que

$$\sum_{x_{n+1}} \min\{\lambda_0 f_0^{n+1}, \lambda_1 f_1^{n+1}\} \leq \sum_{x_{n+1}} \lambda_0 f_0^n f_0(x_{n+1}) = \lambda_0 f_0^n$$

y lo mismo ocurre con  $\lambda_1 f_1^{n+1}$ , por lo que se cumple (2.39).

Para  $n = N$  tenemos que  $V_N^N = m_N$ . Ahora sea  $n = N - 1$ , entonces

$$\begin{aligned}
V_{N-1}^N &= \min\{m_{N-1}, m_{N-1} p_{N-1} + (1 - p_{N-1}) \sum_{x_N} V_N^N\} \\
&= \min\{m_{N-1}, m_{N-1} p_{N-1} + (1 - p_{N-1}) \sum_{x_N} m_N\} \\
&\leq \min\{m_{N-1}, m_{N-1} p_{N-1} + (1 - p_{N-1}) m_{N-1}\} \quad \text{por (2.39)} \\
&= \min\{m_{N-1}, m_{N-1}\} = m_{N-1}.
\end{aligned}$$

Supongamos que para  $n + 1$  se cumple que  $V_{n+1}^N \leq m_{n+1}$ . Probemos que también se cumple para  $n$ .

$$\begin{aligned}
 V_n^N &= \min\{m_n, m_n p_n + (1 - p_n) \sum_{x_{n+1}} V_{n+1}^N\} \\
 &\leq \min\{m_n, m_n p_n + (1 - p_n) \sum_{x_{n+1}} m_{n+1}\} \\
 &\leq \min\{m_n, m_n p_n + (1 - p_n) m_n\} \\
 &= \min\{m_n, m_n\} = m_n.
 \end{aligned}$$

De manera implícita estamos probando que tenemos la regla óptima, toda vez que para toda  $n = N - 1, N - 2, \dots, 1$  se cumple

$$\sum_{x_{n+1}} V_{n+1}^N \leq \sum_{x_{n+1}} m_{n+1} \leq m_n.$$

■

Hasta ahora hemos caracterizado el caso truncado; es decir, cuando  $\psi_N \equiv 1$ , pero queremos tener algo más general. Entonces lo que sigue ahora es estudiar la función de Lagrange  $L_N(\psi)$  y la regla de paro óptima truncada a nivel  $N$  cuando  $N \rightarrow \infty$ .

## 3.2

### Caso no Truncado

En esta parte veremos el caso general de la regla de paro óptima; es decir, daremos la regla de paro óptima en la clase de todas las pruebas secuenciales. Para este propósito haremos uso de lo visto en la sección anterior.

Recordemos que la función de Lagrange truncada al nivel  $N$ , dada cualquier regla de paro  $\psi$ , tiene la expresión dada en (2.23), la cual tiene la forma

$$L_N(\psi) = \sum_{n=1}^{N-1} \sum_{x_1 \cdots x_n} t_n^\psi \left( \psi_n P(D > n) + P(D = n) \right) \left( cn f_0^n + m_n \right)$$

$$+ \sum_{x_1 \cdots x_N} t_N^\psi P(D \geq N) (cN f_0^N + m_N).$$

Entonces si tenemos esta función de Lagrange, con regla de paro truncada al nivel  $N$ , se puede aplicar lo visto en la sección anterior, en particular las desigualdades del Teorema 2.5.

También será útil para esta sección el uso de  $E_{\theta_0}[\nu_\psi \wedge D]$ , pero calculada de la siguiente manera:

$$\begin{aligned} E_{\theta_0}[\nu_\psi \wedge D] &= \sum_{n=1}^{\infty} (\nu_\psi \wedge D) P_{\theta_0}(\nu_\psi \wedge D = n) \\ &= \sum_{n=1}^{\infty} n P_{\theta_0}(\nu_\psi \wedge D = n) \\ &= \sum_{n=1}^{\infty} P_{\theta_0}(\nu_\psi \wedge D \geq n) \\ &= \sum_{n=1}^{\infty} P_{\theta_0}(\nu_\psi \geq n) P(D \geq n). \end{aligned} \tag{2.40}$$

Antes de pasar al objetivo principal de esta parte del trabajo, tenemos los siguientes lemas.

**Lema 2.8.** *Supongamos que  $c > 0$  y sea  $\psi$  cualquier regla de paro, tal que  $E_{\theta_0}[\nu_\psi \wedge D] < \infty$ , entonces*

$$\lim_{N \rightarrow \infty} L_N(\psi; c, \lambda_0, \lambda_1) = L(\psi; c, \lambda_0, \lambda_1).$$

*Demostración.*

Hagamos la siguiente diferencia

$$\begin{aligned} L(\psi; c, \lambda_0, \lambda_1) - L_N(\psi; c, \lambda_0, \lambda_1) &= \sum_{n=1}^{\infty} \sum_{x_1 \cdots x_n} t_n^\psi (\psi_n P(D > n) + P(D = n)) (c n f_0^n + m_n) \\ &\quad - \sum_{n=1}^{N-1} \sum_{x_1 \cdots x_n} t_n^\psi (\psi_n P(D > n) + P(D = n)) (c n f_0^n + m_n) \\ &\quad - \sum_{x_1 \cdots x_N} t_N^\psi P(D \geq N) (c N f_0^N + m_N) \\ &= \sum_{n=N}^{\infty} \sum_{x_1 \cdots x_n} t_n^\psi (\psi_n P(D > n) + P(D = n)) (c n f_0^n + m_n) \end{aligned}$$

$$- \sum_{x_1 \cdots x_N} t_N^\psi P(D \geq N) (cNf_0^N + m_N). \quad (2.41)$$

Por la ecuación (2.6), tenemos que

$$L(\psi; c, \lambda_0, \lambda_1) = cE_{\theta_0}[\nu_\psi \wedge D] + \lambda_0\alpha + \lambda_1\beta = \sum_{n=1}^{\infty} \sum_{x_1 \cdots x_n} t_n^\psi (\psi_n P(D > n) + P(D = n)) (cnf_0^n + m_n),$$

donde  $0 < \lambda_i < \infty$  para  $i = 0, 1$  y  $0 < \alpha, \beta < 1$ . Por hipótesis, tenemos que  $c > 0$  y  $E_{\theta_0}[\nu_\psi \wedge D] < \infty$ . Así tenemos que  $L(\psi; c, \lambda_0, \lambda_1) < \infty$ . Por lo tanto la primera suma de (2.41) cumple con

$$\lim_{N \rightarrow \infty} \sum_{n=N}^{\infty} \sum_{x_1 \cdots x_n} t_n^\psi (\psi_n P(D > n) + P(D = n)) (cnf_0^n + m_n) = 0, \quad (2.42)$$

por ser la cola de una serie convergente. Ahora, de la resta de (2.41) vemos que hay dos sumandos  $(cNf_0^N + m_N)$ . Analicemos el primero cuando  $N \rightarrow \infty$ , entonces

$$\begin{aligned} 0 \leq cNP(D \geq N) \sum_{x_1 \cdots x_N} t_N^\psi f_0^N &= cNP(D \geq N)P_{\theta_0}(\nu_\psi \geq N) \\ &= cNP_{\theta_0}(\nu_\psi \geq N, D \geq N) \\ &= cNP_{\theta_0}(\nu_\psi \wedge D \geq N) \\ &= c \sum_{n=N}^{\infty} NP_{\theta_0}(\nu_\psi \wedge D = n) \\ &\leq c \sum_{n=N}^{\infty} nP_{\theta_0}(\nu_\psi \wedge D = n) \\ &= c \sum_{n=N}^{\infty} P_{\theta_0}(\nu_\psi \wedge D \geq n). \end{aligned} \quad (2.43)$$

Sabemos por (2.40) que  $E_{\theta_0}[\nu_\psi \wedge D] = \sum_{n=1}^{\infty} P_{\theta_0}(\nu_\psi \wedge D \geq n) < \infty$ . Por lo tanto, (2.43) tiende a 0 como  $N \rightarrow \infty$ , ya que  $c > 0$  es una constante.

Ahora analizando la segunda suma, podemos encontrar un  $N \geq \lambda_0$ , con lo que tenemos lo siguiente

$$\begin{aligned} 0 \leq P(D \geq N) \sum_{x_1 \cdots x_N} t_N m_N &\leq \lambda_0 P(D \geq N) \sum_{x_1 \cdots x_N} t_N f_0^N \\ &\leq NP(D \geq N)P_{\theta_0}(\nu_\psi \geq N). \end{aligned} \quad (2.44)$$

Así, podemos ver a (2.44) como un caso particular de (2.43) con  $c = 1$ , por lo tanto también (2.44)  $\rightarrow 0$  como  $N \rightarrow \infty$ . Ahora, la resta de (2.41) tiende a cero como  $N$  tiende a infinito.

De esta manera vemos que  $\lim_{N \rightarrow \infty} (L(\psi; c, \lambda_0, \lambda_1) - L_N(\psi; c, \lambda_0, \lambda_1)) = 0$ . ■

El siguiente lema es un caso particular del lema anterior. Al suponer a  $c = 0$  ya no es necesaria la condición  $E_{\theta_0}[\nu_\psi \wedge D] < \infty$ , como veremos.

**Lema 2.9.** *Sea  $\psi$  cualquier regla de paro, entonces*

$$\lim_{N \rightarrow \infty} L_N(\psi; 0, \lambda_0, \lambda_1) = L(\psi; 0, \lambda_0, \lambda_1). \quad (2.45)$$

*Demostración.*

Por el Lema 2.8, con  $c = 0$ , sabemos que

$$\begin{aligned} L(\psi; 0, \lambda_0, \lambda_1) - L_N(\psi; 0, \lambda_0, \lambda_1) &= \sum_{n=N}^{\infty} \sum_{x_1 \cdots x_n} t_n^\psi m_n \left( \psi_n P(D > n) + P(D = n) \right) \\ &\quad - \sum_{x_1 \cdots x_N} t_N^\psi P(D \geq N) m_N. \end{aligned} \quad (2.46)$$

Por (2.42) y con  $c = 0$ , sabemos que la primer suma de (2.46) tiende a cero cuando  $N \rightarrow \infty$ . Por otra parte, al considerar (2.44), vemos que la resta de (2.46) también tiende a cero como  $N \rightarrow \infty$ . Por lo tanto, se cumple (2.45). ■

Lo siguiente es ver qué comportamiento tienen las funciones  $V_k^N$  definidas en el Teorema 2.5 cuando variamos  $N \rightarrow \infty$ . No olvidemos que las funciones  $V_k^N$  dependen de la muestra  $x_1, \dots, x_k$ , analizada hasta el tiempo  $k$ ; es decir,  $V_k^N = V_k^N(x_1, \dots, x_k)$ . El objetivo es que, una vez construida la muestra de tamaño  $k$ , variemos  $N$ . Esto queda en el siguiente.

**Lema 2.10.** *Para cualquier  $r \geq 1$  y para cualquier  $N \geq r$ ,*

$$V_r^N(x_1, \dots, x_r) \geq V_r^{N+1}(x_1, \dots, x_r). \quad (2.47)$$

*Demostración.*

La prueba es por inducción hacia atrás sobre  $r = N, N - 1, N - 2, \dots, 1$  y partiendo de

$V_N^N = m_N$ . Sea  $r = N$ , entonces por (2.34)

$$V_N^{N+1} = \min \left\{ m_N, p_N m_N + (1 - p_N)(c f_0^N + \sum_{x_{N+1}} V_{N+1}^{N+1}) \right\} \leq m_N = V_N^N.$$

Supongamos que (2.47) se cumple para alguna  $r, N \geq r \geq 1$  entonces

$$\begin{aligned} V_{r-1}^N &= \min \left\{ m_{r-1}, p_{r-1} m_{r-1} + (1 - p_{r-1})(c f_0^{r-1} + \sum_{x_r} V_r^N) \right\} \\ &\geq \min \left\{ m_{r-1}, p_{r-1} m_{r-1} + (1 - p_{r-1})(c f_0^{r-1} + \sum_{x_r} V_r^{N+1}) \right\} = V_{r-1}^{N+1}. \end{aligned}$$

Así tenemos que también se cumple para  $r - 1$ , lo que concluye la demostración. ■

Del Lema 2.10 vemos que para cualquier  $N \geq r \geq 1$  la sucesión  $V_r^N(x_1, \dots, x_r)$  es decreciente al variar  $N$ . Además, por (2.34) sabemos que para toda  $N$  y cada  $1 \leq r \leq N$ ,  $V_r^N(x_1, \dots, x_r)$  es no negativo, por lo que existe su límite, al que denotaremos como

$$V_r(x_1, \dots, x_r) = \lim_{N \rightarrow \infty} V_r^N(x_1, \dots, x_r).$$

Entonces, al hacer  $N \rightarrow \infty$  en todas las igualdades y desigualdades en el Teorema 2.5, tenemos de forma inmediata el siguiente.

**Teorema 2.11.** *Sea  $\psi$  cualquier regla de paro. Entonces para cualquier  $r \geq 1$  se cumplen las siguientes desigualdades*

$$\begin{aligned} L(\psi) &\geq \sum_{n=1}^r \sum_{x_1 \cdots x_n} t_n^\psi \left( \psi_n P(D > n) + P(D = n) \right) (c n f_0^n + m_n) \\ &\quad + \sum_{x_1 \cdots x_{r+1}} t_{r+1}^\psi P(D \geq r + 1) (c(r + 1) f_0^{r+1} + V_{r+1}) \\ &\geq \sum_{n=1}^{r-1} \sum_{x_1 \cdots x_n} t_n^\psi \left( \psi_n P(D > n) + P(D = n) \right) (c n f_0^n + m_n) \\ &\quad + \sum_{x_1 \cdots x_r} t_r^\psi P(D \geq r) (c r f_0^r + V_r). \end{aligned}$$

Además, los  $V_r$  cumple con la siguiente identidad:

$$V_r = \min \left\{ m_r, p_r m_r + (1 - p_r)(c f_0^r + \sum_{x_{r+1}} V_{r+1}) \right\}, \quad (2.48)$$

con  $p_r = \frac{P(D=r)}{P(D \geq r)}$ .

Observemos que en este teorema hace falta la regla de paro, en comparación con el Teorema 2.5, para poder alcanzar la cota inferior, misma que así resultaría óptima. Así que ahora estamos en condiciones de ver uno de los objetivos principales de este trabajo.

**Teorema 2.12.** *Sea  $\psi$  cualquier regla de paro. Entonces*

$$L(\psi) \geq c + \sum_{x_1} V_1(x_1). \quad (2.49)$$

Además, para la regla de paro  $\psi^* = (\psi_1^*, \psi_2^*, \psi_3^*, \dots)$ , se cumple

$$L(\psi^*) = c + \sum_{x_1} V_1(x_1), \quad (2.50)$$

donde

$$\psi_r^* = I_{\left\{ m_r \leq c f_0^r + \sum_{x_{r+1}} V_{r+1} \right\}}, \quad r = 1, 2, 3, \dots \quad (2.51)$$

y  $\psi_r^* \equiv 1$  siempre que  $p_r = 1$ , sin importar  $\psi_{r+1}^*, \psi_{r+2}^*, \dots$

*Demostración.*

Sea  $\psi$  una regla de paro arbitraria. Por el Teorema 2.11 y para cualquier  $r \geq 1$  fijo se cumplen las desigualdades:

$$\begin{aligned} L(\psi) &\geq \sum_{n=1}^r \sum_{x_1 \cdots x_n} t_n^\psi \left( \psi_n P(D > n) + P(D = n) \right) \left( c n f_0^n + m_n \right) \\ &\quad + \sum_{x_1 \cdots x_{r+1}} t_{r+1}^\psi P(D \geq r + 1) \left( c(r + 1) f_0^{r+1} + V_{r+1} \right) \end{aligned} \quad (2.52)$$

$$\begin{aligned} &\geq \sum_{n=1}^{r-1} \sum_{x_1 \cdots x_n} t_n^\psi \left( \psi_n P(D > n) + P(D = n) \right) \left( c n f_0^n + m_n \right) \\ &\quad + \sum_{x_1 \cdots x_r} t_r^\psi P(D \geq r) \left( c r f_0^r + V_r \right) \end{aligned} \quad (2.53)$$

$$\begin{aligned}
 &\geq \dots \\
 &\geq \sum_{x_1} t_1^\psi(x_1) \left( \psi_1 P(D > 1) + P(D = 1) \right) \left( c f_0^1(x_1) + m_1(x_1) \right) \\
 &\quad + \sum_{x_1, x_2} t_2^\psi(x_1, x_2) P(D \geq 2) \left( 2c f_0^2(x_1, x_2) + V_2(x_1, x_2) \right) \tag{2.54}
 \end{aligned}$$

$$\geq \sum_{x_1} (c f_0^1(x_1) + V_1(x_1)) = c \sum_{x_1} f_0^1(x_1) + \sum_{x_1} V_1(x_1). \tag{2.55}$$

Como  $f_0^1$  es la función de probabilidad bajo la hipótesis nula de  $x_1$ , se cumple  $\sum_{x_1} f_0^1(x_1) = 1$ . Por lo que hemos probado (2.49).

Ahora queremos ver cuándo tenemos igualdades en (2.52)-(2.55). La idea es analizar de la cota inferior (2.55) hacia arriba. Para este propósito haremos uso del Lema 2.4. Si tomamos a  $r = 2$ ,  $v_1 = V_1$ ,  $v_2 = V_2$ , y los  $V_i$  cumplen con (2.48), además si

$$\psi_1 = I_{\left\{ m_1 \leq c f_0^1 + \sum_{x_2} V_2(x_1, x_2) \right\}} = \psi_1^*,$$

entonces al aplicar el Lema 2.4 podemos ver que tendremos igualdad en (2.55). Pero la igualdad se da también si  $\psi_1 \equiv 1$  siempre que  $p_1 = 1$  y todas las desigualdades se vuelven igualdades, porque  $P(D > 1) = 0$  y terminamos.

Ahora si  $p_1 < 1$ ,  $r = 3$ ,  $v_2 = V_2$ ,  $v_3 = V_3$  y  $V_i$  cumple con la identidad (2.48)  $i = 2, 3$ , además si

$$\psi_2 = I_{\left\{ m_2 \leq c f_0^2 + \sum_{x_3} V_3 \right\}} = \psi_2^*,$$

entonces al aplicar nuevamente el Lema 2.4, tendremos ahora igualdad en (2.54). Pero la igualdad se da también si  $\psi_2 \equiv 1$  siempre que  $p_2 = 1$  y todas las desigualdades restantes se vuelven igualdades, porque  $P(D > 2) = 0$  y terminamos.

Podemos aplicar el Lema 2.4 de manera sucesiva, mientras  $p_i < 1$   $i = 2, 3, \dots, r$ ; tomando los  $v_i = V_i$  y los  $V_i$  que cumplan con la identidad dada en (2.48) y  $\psi_i = \psi_i^*$  ( $\psi_i^*$  como en (2.51)) para  $i = 3, \dots, r + 1$ , tendremos que todas las desigualdades, excepto quizás la primera, se vuelven igualdades para la regla

$$\psi^* = (\psi_1^*, \psi_2^*, \psi_3^*, \dots, \psi_r^*, \dots).$$

Si  $p_i = 1$  para alguna  $2 \leq i \leq r$  (de hecho ya sabemos que es única), entonces la componente  $i$  de la regla de paro cambia a  $\psi_i^* \equiv 1$  y las desigualdades restantes se vuelven igualdades, con lo que terminamos.

Entonces para la regla  $\psi^*$  (quizás con  $\psi_i^* \equiv 1$  para alguna  $i$ , sin importar  $\psi_{i+1}^*, \psi_{i+2}^*, \dots$ ), se tiene que

$$\begin{aligned} & \sum_{n=1}^r \sum_{x_1 \cdots x_n} t_n^{\psi^*} \left( \psi_n^* P(D > n) + P(D = n) \right) (cnf_0^n + m_n) \\ & + \sum_{x_1 \cdots x_{r+1}} t_{r+1}^{\psi^*} P(D \geq r+1) (c(r+1)f_0^{r+1} + V_{r+1}) = c + \sum_{x_1} V_1(x_1), \end{aligned} \quad (2.56)$$

para toda  $r \in \mathbb{N}$ .

De esto último se tiene la siguiente desigualdad

$$\sum_{n=1}^r \sum_{x_1 \cdots x_n} t_n^{\psi^*} \left( \psi_n^* P(D > n) + P(D = n) \right) (cnf_0^n + m_n) \leq c + \sum_{x_1} V_1(x_1). \quad (2.57)$$

Probemos que la parte izquierda de (2.57) es  $L(\psi^*)$  cuando  $r \rightarrow \infty$ . Esto último es cierto si  $P_{\theta_0}(\nu_{\psi^*} \wedge D < \infty) = 1$  o si  $P_{\theta_0}(\nu_{\psi^*} \wedge D = \infty) = 0$ , que es su complemento.

Observemos que de (2.56) se cumple lo siguiente.

$$\begin{aligned} c + \sum_{x_1} V_1(x_1) & \geq \sum_{x_1 \cdots x_{r+1}} t_{r+1}^{\psi^*} P(D \geq r+1) (c(r+1)f_0^{r+1} + V_{r+1}) \\ & \geq \sum_{x_1 \cdots x_{r+1}} t_{r+1}^{\psi^*} P(D \geq r+1) (c(r+1)f_0^{r+1}) \\ & = c(r+1)P(D \geq r+1) \sum_{x_1 \cdots x_{r+1}} t_{r+1}^{\psi^*} f_0^{r+1} \\ & = c(r+1)P(D \geq r+1)P_{\theta_0}(\nu_{\psi^*} \geq r+1). \end{aligned}$$

Entonces

$$cP(D \geq r+1)P_{\theta_0}(\nu_{\psi^*} \geq r+1) \leq \frac{c + \sum_{x_1} V_1(x_1)}{r+1},$$

$$cP_{\theta_0}(\nu_{\psi^*} \wedge D \geq r+1) \leq \frac{c + \sum_{x_1} V_1(x_1)}{r+1}. \quad (2.58)$$

Si tomamos el límite  $r \rightarrow \infty$  en (2.58), tenemos que  $P_{\theta_0}(\nu_{\psi^*} \wedge D = \infty) = 0$ , puesto que  $c + \sum_{x_1} V_1(x_1) = c + \sum_{x_1} \min\{\lambda_0 f_{\theta_0}(x_1), \lambda_1 f_{\theta_1}(x_1)\}$  es un número real positivo. Entonces al hacer  $r \rightarrow \infty$  en (2.57), se cumple que

$$L(\psi^*) \leq c + \sum_{x_1} V_1(x_1). \quad (2.59)$$

Por otra parte y en virtud de (2.49), con  $\psi^*$  como regla de paro, se tiene

$$L(\psi^*) \geq c + \sum_{x_1} V_1(x_1). \quad (2.60)$$

Por lo tanto, de (2.59) y (2.60) tenemos (2.50) ■

Con este teorema podemos ver la forma que tiene la regla  $\psi^*$  que nos ayuda a detener el experimento secuencial. Pero además, hay que recordar que dicha regla de paro va acompañada de una regla de decisión  $\delta^*$ , la cual está dada en (2.13). Por lo que las reglas de decisión  $\delta^*$  y de paro  $\psi^*$  las incorporamos a la función de Lagrange (2.22). Ahora bien, por los teoremas 2.1 y 2.2, podemos observar que hemos obtenido la forma de la prueba secuencial que minimiza el número promedio de observaciones en la clase de todas las pruebas  $\Delta(\alpha, \beta)$  ( $0 < \alpha, \beta < 1$ ), cuyas probabilidades de error no exceden las cotas  $\alpha$  y  $\beta$ .

Observemos que por (2.60),  $C(\psi^*) = cE_{\theta_0}[\nu_{\psi^*} \wedge D] < \infty$ . Por lo tanto, la función de costo  $C(\psi^*)$  se minimiza con la regla de paro  $\psi^*$ .

En otras palabras, hemos hallado la prueba óptima  $(\psi^*, \delta^*) \in \Delta(\alpha, \beta)$ ; es decir,

$$L(\psi^*) = \inf_{\psi} \{cE_{\theta_0}[\nu_{\psi} \wedge D] + \lambda_0\alpha(\psi, \delta^*) + \lambda_1\beta(\psi, \delta^*)\} = c + \sum_{x_1} V_1(x_1). \quad (2.61)$$

Debido a que hemos alcanzado la cota inferior, tenemos que  $\inf = \min$ .

Veamos qué sucede cuando  $c = 0$ , pero en este caso con el horizonte no acotado. Este resultado es de suma importancia, ya que lo utilizaremos en el siguiente capítulo.

**Corolario 2.13.** *Sea  $\psi$  cualquier regla de paro y sea  $p_n < 1$  para toda  $n = 1, 2, 3, \dots$ . Entonces*

$$L(\psi; 0, \lambda_0, \lambda_1) \geq \sum_{n=1}^{\infty} \left( \lambda_0 P_{\theta_0}(\lambda_0 f_0^n \leq \lambda_1 f_1^n) + \lambda_1 P_{\theta_1}(\lambda_0 f_0^n > \lambda_1 f_1^n) \right) P(D = n). \quad (2.62)$$

La cota (2.62) se alcanza cuando la regla de paro es  $\psi^* = (\psi_1^* \equiv 0, \psi_2^* \equiv 0, \psi_3^* \equiv 0, \dots)$ .

*Demostración.*

Veamos primero el siguiente lema que nos servirá en la demostración.

**Lema 2.14.** Sean  $X_1, X_2, \dots, X_n$  observaciones consecutivas, independientes e idénticamente distribuidas (i.i.d) y sea  $Z_n = \prod_{i=1}^n \frac{f_{\theta_1}(x_i)}{f_{\theta_0}(x_i)}$ , con  $f_{\theta_j}(x_j) > 0$  para  $j = 1, 2$  y para toda  $i = 1, 2, \dots, n$ . Entonces dado  $\epsilon > 0$ , se cumple que

$$\lim_{n \rightarrow \infty} P_{\theta_0}(Z_n \geq \epsilon) = 0 \quad (2.63)$$

y

$$\lim_{n \rightarrow \infty} P_{\theta_1}(Z_n < \epsilon) = 0. \quad (2.64)$$

*Demostración.*

Iniciemos la demostración para la probabilidad bajo  $\theta_0$ . Por la desigualdad de Chebyshev se cumple

$$\begin{aligned} P_{\theta_0}(Z_n \geq \epsilon) &= P_{\theta_0}(Z_n^{1/2} \geq \epsilon^{1/2}) \\ &\leq \epsilon^{-1/2} E_{\theta_0}[Z_n^{1/2}] \\ &= \epsilon^{-1/2} E_{\theta_0} \left[ \left( \prod_{i=1}^n \frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)} \right)^{1/2} \right] \\ &= \epsilon^{-1/2} E_{\theta_0} \left[ \left( \frac{f_{\theta_1}(X_1)f_{\theta_1}(X_2) \cdots f_{\theta_1}(X_n)}{f_{\theta_0}(X_1)f_{\theta_0}(X_2) \cdots f_{\theta_0}(X_n)} \right)^{1/2} \right] \\ &= \epsilon^{-1/2} \left( E_{\theta_0} \left[ \left( \frac{f_{\theta_1}(X)}{f_{\theta_0}(X)} \right)^{1/2} \right] \right)^n \quad \text{por ser } X_i \text{ i.i.d.} \end{aligned}$$

Si probamos que  $E_{\theta_0} \left[ \left( \frac{f_{\theta_1}(X)}{f_{\theta_0}(X)} \right)^{1/2} \right] < 1$ , entonces habremos terminado.

Por la desigualdad de Cauchy-Schwarz  $(E[XY])^2 \leq E[X^2]E[Y^2]$  con  $X, Y$  como variables aleatorias), haciendo  $X = \frac{f_{\theta_1}(X)}{f_{\theta_0}(X)}$  y  $Y = 1$ , tenemos que

$$\begin{aligned} E_{\theta_0} \left[ \left( \frac{f_{\theta_1}(X)}{f_{\theta_0}(X)} \right)^{1/2} \right] &\leq E_{\theta_0} \left[ \frac{f_{\theta_1}(X)}{f_{\theta_0}(X)} \right]^{1/2} \\ &= \left( \sum_x \frac{f_{\theta_1}(x)}{f_{\theta_0}(x)} f_{\theta_0}(x) \right)^{1/2} = \left( \sum_x f_{\theta_1}(x) \right)^{1/2} = 1. \end{aligned}$$

Sabemos que la igualdad sólo se da cuando  $f_{\theta_1}(x) = \lambda f_{\theta_0}(x)$ , para toda  $x$  con probabilidad 1, para alguna  $\lambda \in \mathbb{R}$ . Luego,  $\lambda > 0$  porque  $f_{\theta_j}(x) > 0$  con  $j = 1, 2$ . Así,  $\frac{f_{\theta_1}(x)}{f_{\theta_0}(x)} = \lambda$  para toda  $x$  con probabilidad 1. Entonces,

$$E[\lambda]^{1/2} = \lambda^{1/2} = 1 \quad \text{con probabilidad 1.}$$

Por lo tanto,  $E_{\theta_0} \left[ \left( \frac{f_{\theta_1}(X)}{f_{\theta_0}(X)} \right)^{1/2} \right] < 1$ , con lo que hemos probado (2.63).

Falta por demostrar (2.64).

$$\begin{aligned} P_{\theta_1}(Z_n < \epsilon) &= P_{\theta_1}(Z_n^{-1/2} > \epsilon^{-1/2}) \leq \epsilon^{1/2} E_{\theta_1}[Z_n^{-1/2}] \\ &= \epsilon^{1/2} E_{\theta_1} \left[ \left( \prod_{i=1}^n \frac{f_{\theta_0}(X_i)}{f_{\theta_1}(X_i)} \right)^{1/2} \right] \\ &= \epsilon^{1/2} \left( E_{\theta_1} \left[ \left( \frac{f_{\theta_0}(X)}{f_{\theta_1}(X)} \right)^{1/2} \right] \right)^n \end{aligned} \quad (2.65)$$

Como  $E_{\theta_1} \left[ \left( \frac{f_{\theta_0}(X)}{f_{\theta_1}(X)} \right)^{1/2} \right] \leq 1$ , procedemos como en el caso de  $\theta_0$  y se cumple (2.64) al tomar el límite cuando  $n \rightarrow \infty$  en (2.65), con lo que concluimos la demostración de este lema. ■

Regresando con la prueba de (2.62), sea  $\psi$  cualquier regla de paro truncada a nivel  $N$  y haciendo uso del Corolario 2.7 se sigue que

$$\begin{aligned} L_N(\psi; 0, \lambda_0, \lambda_1) &\geq \lambda_0 P_{\theta_0}(\lambda_0 f_0^N \leq \lambda_1 f_1^N) P(D > N) + \lambda_1 P_{\theta_1}(\lambda_0 f_0^N > \lambda_1 f_1^N) P(D > N) \\ &\quad + \lambda_0 \sum_{n=1}^N P_{\theta_0}(\lambda_0 f_0^n \leq \lambda_1 f_1^n) P(D = n) + \lambda_1 \sum_{n=1}^N P_{\theta_1}(\lambda_0 f_0^n > \lambda_1 f_1^n) P(D = n) \\ &= \lambda_0 P_{\theta_0}(\lambda_0/\lambda_1 \leq f_1^N/f_0^N) P(D > N) + \lambda_1 P_{\theta_1}(\lambda_0/\lambda_1 > f_1^N/f_0^N) P(D > N) \\ &\quad + \lambda_0 \sum_{n=1}^N P_{\theta_0}(\lambda_0 f_0^n \leq \lambda_1 f_1^n) P(D = n) + \lambda_1 \sum_{n=1}^N P_{\theta_1}(\lambda_0 f_0^n > \lambda_1 f_1^n) P(D = n) \end{aligned}$$

Al tomar el límite cuando  $N \rightarrow \infty$  en ambos miembros de la desigualdad anterior y tomando en cuenta que por el Lema 2.14, tenemos que  $\lim_{N \rightarrow \infty} P_{\theta_0}(\lambda_0/\lambda_1 \leq f_1^N/f_0^N) = 0$  y  $\lim_{N \rightarrow \infty} P_{\theta_1}(\lambda_0/\lambda_1 > f_1^N/f_0^N) = 0$ . Además, como  $c = 0$  podemos hacer uso del Lema 2.8, entonces llegamos a (2.62).

Al ser  $c = 0$ , lo conveniente es seguir observando los datos ya que no hay costo alguno y obtenemos más información. Por lo tanto, la regla de paro óptima es  $\psi^* = (\psi_1^* \equiv 0, \psi_2^* \equiv 0, \psi_3^* \equiv 0, \dots)$  y por el Teorema 2.12 se alcanza la cota (2.62). ■

Lo que sigue es ver la estructura completa de la prueba secuencial óptima.

## 4

### Estructura de la Prueba Secuencial Óptima

En esta última sección mostraremos que el objetivo principal de este capítulo es alcanzado al obtener la optimalidad de la prueba secuencial con horizonte aleatorio, tal como veremos en el siguiente.

**Teorema 2.15.** *Sea  $(\psi^*, \delta^*)$  una prueba secuencial, con  $\delta^* = (\delta_1^*, \delta_2^*, \delta_3^*, \dots)$ , donde*

$$\delta_n^* = I_{\{\lambda_0 f_0^n - \lambda_1 f_1^n \leq 0\}} = \begin{cases} 1 & \text{si } \lambda_0 f_0^n - \lambda_1 f_1^n \leq 0, \\ 0 & \text{en otro caso,} \end{cases} \quad (2.66)$$

y  $\psi^* = (\psi_1^*, \psi_2^*, \psi_3^*, \dots)$ , con

$$\psi_r^* = I_{\{m_r \leq c f_0^r + \sum_{x_{r+1}} V_{r+1}\}}, \quad r = 1, 2, 3, \dots \quad (2.67)$$

y los  $V_{r+1} = V_{r+1}(x_1, \dots, x_{r+1})$  cumplen con (2.48).

Entonces para cualquier otra prueba secuencial  $(\psi, \delta)$ , tal que

$$\alpha(\psi, \delta) \leq \alpha(\psi^*, \delta^*) \quad \text{y} \quad \beta(\psi, \delta) \leq \beta(\psi^*, \delta^*), \quad (2.68)$$

se cumple

$$E_{\theta_0}[\nu_{\psi^*} \wedge D] \leq E_{\theta_0}[\nu_{\psi} \wedge D]. \quad (2.69)$$

En caso de que el horizonte  $D$  sea acotado, ( $p_k = 1$  para algún  $k \in \mathbb{N}$ ), entonces únicamente cambia la componente  $k$ -ésima de la regla (2.67) a  $\psi_k^* \equiv 1$ , sin importar la definición de  $\psi_{k+1}, \psi_{k+2}, \dots$  y en este caso  $V_k = m_k$ .

La desigualdad en (2.69) es estricta si alguna de las desigualdades en (2.68) es estricta.

*Demostración.*

Sea  $(\psi, \delta)$  cualquier prueba secuencial. Por el Teorema 2.2

$$L(\psi; c, \lambda_0, \lambda_1) = L(\psi, \delta^*; c, \lambda_0, \lambda_1) \leq L(\psi, \delta; c, \lambda_0, \lambda_1). \quad (2.70)$$

A su vez, por el Teorema 2.12

$$L(\psi^*; c, \lambda_0, \lambda_1) \leq L(\psi; c, \lambda_0, \lambda_1),$$

así que en vista de (2.70) se tiene

$$L(\psi^*, \delta^*; c, \lambda_0, \lambda_1) = L(\psi^*; c, \lambda_0, \lambda_1) \leq L(\psi; c, \lambda_0, \lambda_1) \leq L(\psi, \delta; c, \lambda_0, \lambda_1),$$

lo que significa que se cumple la condición (2.7) del Teorema 2.1.

Denotemos  $\alpha = \alpha(\psi^*, \delta^*)$  y  $\beta = \beta(\psi^*, \delta^*)$ . Entonces para cualquier prueba que satisfice (2.68) se cumple

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

y por el Teorema 2.1 se sigue que

$$C(\psi^*) = cE_{\theta_0}[\nu_{\psi^*} \wedge D] \leq C(\psi) = cE_{\theta_0}[\nu_{\psi} \wedge D],$$

que es equivalente a (2.69) ya que  $c > 0$ . Este resultado es válido para la regla  $\psi^*$  ya sea que el horizonte  $D$  sea acotado o no.

Además, por el mismo Teorema 2.1, la desigualdad en (2.69) es estricta si alguna de las desigualdades de (2.68) es estricta. ■

Este teorema nos dice que con la prueba  $(\psi^*, \delta^*)$  obtenemos la optimalidad de la prueba secuencial estadística; esto es, minimizamos el número promedio de observaciones en la clase de todas las pruebas cuyas probabilidades de error no exceden las cotas (2.68). Esto significa minimizar la función de costo del experimento, ya sea con horizonte aleatorio  $D$  acotado o no, relacionado con la función de riesgo  $p_k$  o también conocida como tasa de fallos.

El mínimo de observaciones lo calculamos bajo la hipótesis nula  $H_0$ , sin embargo, para la misma prueba, se puede demostrar que también se minimiza el número de observaciones bajo la hipótesis alternativa  $H_1$  [12].

En este capítulo la independencia de las observaciones no se usa de manera esencial por lo que todos los resultados pueden ser aplicados a cualquier proceso estocástico a tiempo discreto.

## CAPÍTULO 3

# PRUEBAS ÓPTIMAS Y RAZÓN DE PROBABILIDADES

## 1

### Estructura de Pruebas Óptimas en Términos de Razón de Probabilidades

En esta sección plantearemos la prueba óptima, obtenida en la última sección del capítulo anterior, en la forma de la razón de probabilidades en caso de horizonte aleatorio. Estudiaremos las variantes que hay con la famosa prueba de Wald la llamada *Prueba Secuencial de Razón de Probabilidades* (SPRT). Es natural ver qué diferencias existen, ya que en la teoría clásica, como mencionamos anteriormente, se considera un horizonte infinito y no uno aleatorio que puede ser acotado.

Para esta parte haremos uso de las funciones  $V_r^N = V_r^N(x_1, \dots, x_r)$  dadas en (2.34). También es conveniente recordar que  $m_r = \min\{\lambda_0 f_0^r, \lambda_1 f_1^r\}$ . Además, para evitar complicaciones técnicas, supondremos que para todos los valores  $x$ , se cumple

$$f_{\theta_0}(x) > 0 \quad \text{y} \quad f_{\theta_1}(x) > 0.$$

Definamos como

$$Z_n = Z_n(x_1, \dots, x_n) = \prod_{i=1}^n \frac{f_{\theta_1}(x_i)}{f_{\theta_0}(x_i)} = \frac{f_1^n}{f_0^n}, \quad (3.1)$$

que es conocida como la *razón de probabilidades*, con  $n = 1, 2, \dots$

Para expresar los elementos de la regla de paro óptima en términos de  $Z_n$  es conveniente definir unas nuevas funciones  $\rho_r^N : \mathbb{R}^+ \rightarrow \mathbb{R}$  como sigue.

Para  $r = N$

$$\rho_N^N(z) = \min\{\lambda_0, \lambda_1 z\} = g(z) \quad \forall z \in \mathbb{R}^+. \quad (3.2)$$

y para  $r < N$

$$\rho_r^N(z) = \min \left\{ g(z), g(z)p_r + (1 - p_r) \left( c + E_{\theta_0} \left[ \rho_{r+1}^N \left( z \frac{f_1(X)}{f_0(X)} \right) \right] \right) \right\} \quad \forall z \in \mathbb{R}^+. \quad (3.3)$$

Estas nuevas funciones guardan cierta relación con las funciones  $V_r^N$ , tal como lo muestra el siguiente.

**Lema 3.1.** *Sea  $\rho_r^N$  definida como en (3.2) y (3.3). Entonces para  $Z_r$  dada en (3.1),*

$$f_0^r \rho_r^N(Z_r) = V_r^N, \quad (3.4)$$

para toda  $1 \leq r \leq N$ .

*Demostración.*

Procedamos por inducción inversa. Entonces para  $r = N$  tenemos que

$$f_0^N \rho_N^N(Z_N) = f_0^N g(Z_N) = f_0^N \min\{\lambda_0, \lambda_1 Z_N\} = \min \left\{ f_0^N \lambda_0, \lambda_1 f_0^N \frac{f_1^N}{f_0^N} \right\} = V_N^N.$$

Ahora supongamos que para  $r + 1$  se cumple la igualdad

$$f_0^{r+1} \rho_{r+1}^N(Z_{r+1}) = V_{r+1}^N.$$

Entonces de (3.3), para  $r < N$  y tomando en cuenta que

$$E_{\theta_0} \left[ \rho_{r+1}^N \left( z \frac{f_1(X)}{f_0(X)} \right) \right] = \sum_{x_{r+1}} f_0(x_{r+1}) \rho_{r+1}^N \left( Z_r \frac{f_1(x_{r+1})}{f_0(x_{r+1})} \right),$$

tenemos que

$$\begin{aligned}
 f_0^r \rho_r^N(Z_r) &= \min \left\{ f_0^r g(Z_r), f_0^r g(Z_r) p_r + f_0^r (1 - p_r) \left( c + \sum_{x_{r+1}} f_0(x_{r+1}) \rho_{r+1}^N \left( Z_r \frac{f_1(x_{r+1})}{f_0(x_{r+1})} \right) \right) \right\} \\
 &= \min \left\{ m_r, m_r p_r + (1 - p_r) \left( c f_0^r + \sum_{x_{r+1}} f_0^r f_0(x_{r+1}) \rho_{r+1}^N \left( \frac{f_1^r f_1(x_{r+1})}{f_0^r f_0(x_{r+1})} \right) \right) \right\} \\
 &= \min \left\{ m_r, m_r p_r + (1 - p_r) \left( c f_0^r + \sum_{x_{r+1}} f_0^{r+1} \rho_{r+1}^N(Z_{r+1}) \right) \right\} \\
 &= \min \left\{ m_r, m_r p_r + (1 - p_r) \left( c f_0^r + \sum_{x_{r+1}} V_{r+1}^N \right) \right\} \\
 &= V_r^N.
 \end{aligned}$$

Por lo tanto, también se cumple (3.4) para  $r$  en el punto  $Z_r$ . ■

Por el anterior lema podemos tratar a las funciones  $V_r^N$  por medio de las funciones  $f_0^r \rho_r^N$  en los puntos  $Z_r$ .

Como  $f_0^r > 0$  para toda  $r \geq 1$ , tenemos que en los puntos  $Z_r$ ,  $\rho_r^N$  hereda las propiedades de  $V_r^N$ ; es decir,  $\rho_r^N(Z_r) \geq \rho_r^{N+1}(Z_r) \geq 0$ . Entonces para todo  $Z_r$  fijo podemos hablar del límite cuando  $N$  tiende a infinito. Denotemos a este límite como

$$f_0^r \rho_r(Z_r) = \lim_{N \rightarrow \infty} f_0^r \rho_r^N(Z_r) = V_r. \quad (3.5)$$

Veamos que la propiedad anterior se cumple también para toda  $z \in \mathbb{R}^+$ . El siguiente lema nos muestra que las funciones  $\rho_r^N$  son decrecientes cuando  $N \rightarrow \infty$ .

**Lema 3.2.** *Para todo  $z \in \mathbb{R}^+$ , se cumple*

$$\rho_r^N(z) \geq \rho_r^{N+1}(z). \quad (3.6)$$

*Demostración.*

Nuevamente utilizaremos inducción hacia atrás sobre  $r$ . Sea  $r = N$ , entonces por (3.3)

$$\rho_N^{N+1}(z) = \min \left\{ g(z), g(z) p_N + (1 - p_N) \left( c + E_{\theta_0} \left[ \rho_{N+1}^N \left( z \frac{f_1(X)}{f_0(X)} \right) \right] \right) \right\} \leq g(z) = \rho_N^N(z),$$

para toda  $z \in \mathbb{R}^+$ . Esto es cierto por (3.2).

Ahora supongamos que (3.6) se cumple para alguna  $r, N \geq r \geq 1$  y para toda  $z \in \mathbb{R}^+$ . Así vemos que si  $\rho_r^N\left(z \frac{f_1(X)}{f_0(X)}\right) \geq \rho_r^{N+1}\left(z \frac{f_1(X)}{f_0(X)}\right)$ , entonces por propiedad de la esperanza, tenemos

$$E_{\theta_0}\left[\rho_r^N\left(z \frac{f_1(X)}{f_0(X)}\right)\right] \geq E_{\theta_0}\left[\rho_r^{N+1}\left(z \frac{f_1(X)}{f_0(X)}\right)\right] \quad \forall z \in \mathbb{R}^+.$$

Luego,

$$\begin{aligned} \rho_{r-1}^N(z) &= \min \left\{ g(z), g(z)p_{r-1} + (1 - p_{r-1})\left(c + E_{\theta_0}\left[\rho_r^N\left(z \frac{f_1(X)}{f_0(X)}\right)\right]\right) \right\} \\ &\geq \min \left\{ g(z), g(z)p_{r-1} + (1 - p_{r-1})\left(c + E_{\theta_0}\left[\rho_r^{N+1}\left(z \frac{f_1(X)}{f_0(X)}\right)\right]\right) \right\} = \rho_{r-1}^{N+1}(z), \end{aligned}$$

para toda  $z \in \mathbb{R}^+$ . Así vemos que también se cumple para  $r - 1$  y para toda  $z \in \mathbb{R}^+$ . ■

Como  $\{\rho_r^n\}_n$  y  $\left\{E_{\theta_0}\left[\rho_r^n\left(z \frac{f_1(X)}{f_0(X)}\right)\right]\right\}_n$  son sucesiones de números decrecientes no negativos, entonces existe el límite en ambas sucesiones. Denotemos a

$$\rho_r(z) = \lim_{n \rightarrow \infty} \rho_r^n(z) \quad \forall z \in \mathbb{R}^+.$$

De lo anterior se sigue que

$$E_{\theta_0}\left[\rho_r\left(z \frac{f_1(X)}{f_0(X)}\right)\right] = \lim_{n \rightarrow \infty} E_{\theta_0}\left[\rho_r^n\left(z \frac{f_1(X)}{f_0(X)}\right)\right] \quad \forall z \in \mathbb{R}^+.$$

Entonces pasando al límite en (3.3), cuando  $n \rightarrow \infty$ , tenemos que para la función  $\rho_r$  se cumple la siguiente igualdad:

$$\rho_r(z) = \min \left\{ g(z), g(z)p_r + (1 - p_r)\left(c + E_{\theta_0}\left[\rho_{r+1}\left(z \frac{f_1(X)}{f_0(X)}\right)\right]\right) \right\}. \quad (3.7)$$

Como en los puntos  $Z_r, f_0^r \rho_r(Z_r)$  y  $V_r$  coinciden, entonces lo que sigue es expresar la forma de la prueba secuencial óptima  $(\psi^*, \delta^*)$  del Teorema 2.15 en términos de la razón de probabilidades  $Z_r$ , así como  $V_{r+1}$  y todos los demás elementos de la fórmula. Entonces para  $r = 1, 2, 3, \dots$  y tomando en cuenta (3.5), tenemos que

$$\psi_r^* = I_{\left\{m_r \leq cf_0^r + \sum_{x_{r+1}} V_{r+1}\right\}}$$

$$\begin{aligned}
 &= I_{\{f_0^r g(Z_r) \leq c f_0^r + \sum_{x_{r+1}} f_0^{r+1} \rho_{r+1}(Z_{r+1})\}} \\
 &= I_{\left\{f_0^r \left(g(Z_r) \leq c + \sum_{x_{r+1}} f_{\theta_0}(x_{r+1}) \rho_{r+1} \left(Z_r \frac{f_1(x_{r+1})}{f_0(x_{r+1})}\right)\right)\right\}}.
 \end{aligned}$$

Como  $f_0^r > 0$ , entonces la regla de paro  $\psi_r^*$  equivale finalmente a

$$\psi_r^* = I_{\left\{g(Z_r) \leq c + \sum_{x_{r+1}} f_{\theta_0}(x_{r+1}) \rho_{r+1} \left(Z_r \frac{f_1(x_{r+1})}{f_0(x_{r+1})}\right)\right\}}. \quad (3.8)$$

Por otra parte, la regla de decisión  $\delta_r^*$  de (2.66) es simplemente

$$\delta_r^* = I_{\{\lambda_0 \leq \lambda_1 Z_r\}}.$$

El resultado anterior ya es más explícito que el mostrado en (2.51), pero se le puede dar una forma más específica. Denotemos a

$$h_r^n(z) = E_{\theta_0} \left[ \rho_{r+1}^n \left( z \frac{f_1(X_{r+1})}{f_0(X_{r+1})} \right) \right] = \sum_{x_{r+1}} f_{\theta_0}(x_{r+1}) \rho_{r+1}^n \left( z \frac{f_1(x_{r+1})}{f_0(x_{r+1})} \right)$$

y

$$h_r(z) = E_{\theta_0} \left[ \rho_{r+1} \left( z \frac{f_1(X_{r+1})}{f_0(X_{r+1})} \right) \right] = \sum_{x_{r+1}} f_{\theta_0}(x_{r+1}) \rho_{r+1} \left( z \frac{f_1(x_{r+1})}{f_0(x_{r+1})} \right). \quad (3.9)$$

Así veremos que la desigualdad que define la regla de paro en (3.8):

$$g(z) \leq c + h_r(z)$$

equivale simplemente a

$$z \notin (A_r, B_r),$$

con ciertas constantes  $A_r$  y  $B_r$ , tales que  $0 < A_r < B_r < \infty$ , por lo tanto (3.8) toma la forma

$$\psi_r^* = I_{\{Z_r \notin (A_r, B_r)\}}. \quad (3.10)$$

Como hemos estado suponiendo que estamos ante el caso no trivial, es decir,  $\lambda_0 > 0$  y  $\lambda_1 > 0$ , entonces la regla de decisión la podemos poner de manera equivalente como:

$$\delta_r^* = I_{\{\lambda_0/\lambda_1 \leq Z_r\}}.$$

Para verificar la equivalencia entre (3.8) y (3.10) trataremos las propiedades de las funciones  $\rho_r^n$ ,  $\rho_r$ ,  $h_r^n$  y  $h_r$  en el siguiente.

**Lema 3.3.** Las funciones  $\rho_r^n(z)$  definidas en (3.2)-(3.3) y la función  $\rho_r(z)$  de (3.7), así como  $h_r^n$  y  $h_r$  poseen las siguientes propiedades:

- i) son cóncavas y continuas en  $[0, \infty)$ ,
- ii) son no decrecientes en  $[0, \infty)$ ,
- iii)  $\rho_r^n(0) = \rho_r(0) = h_r^n(0) = h_r(0) = 0$  y  $\rho_r^n(\infty) = \rho_r(\infty) = h_r^n(\infty) = h_r(\infty) = \lambda_0$ .

*Demostración.*

En los tres puntos aplicaremos inducción hacia atrás empezando por  $r = n, n - 1, \dots, 1$ .

- i) Para  $r = n$  tenemos que  $\rho_n^n(z) = g(z) = \min\{\lambda_0, \lambda_1 z\}$  ( $\lambda_0, \lambda_1 \geq 0$ ) es cóncava por ser el mínimo de dos funciones cóncavas, más aún como mínimo de funciones lineales. Ahora supongamos que para  $r + 1$ ,  $\rho_{r+1}^n(z)$  es cóncava en  $[0, \infty)$ . Entonces por (3.3) basta probar que  $h_r^n(z)$  es cóncava  $\forall z \in \mathbb{R}^+$  y de esta manera  $\rho_r^n$  es mínimo entre una función cóncava y una suma de funciones cóncavas, ya que  $p_r \in [0, 1]$ , por lo que tendremos que  $\rho_r^n$  también es cóncava.

Sean  $x, y \in \mathbb{R}^+$  arbitrarios. Queremos demostrar que

$$h_r^n(\gamma x + (1 - \gamma)y) \geq \gamma h_r^n(x) + (1 - \gamma)h_r^n(y) \quad \forall \gamma \in [0, 1].$$

Entonces dado  $\gamma \in [0, 1]$  tenemos que

$$\begin{aligned} h_r^n(\gamma x + (1 - \gamma)y) &= E_{\theta_0} \left[ \rho_{r+1}^n \left( (\gamma x + (1 - \gamma)y) \frac{f_1(X)}{f_0(X)} \right) \right] \\ &= E_{\theta_0} \left[ \rho_{r+1}^n \left( \gamma x \frac{f_1(X)}{f_0(X)} + (1 - \gamma)y \frac{f_1(X)}{f_0(X)} \right) \right] \\ &\geq E_{\theta_0} \left[ \gamma \rho_{r+1}^n \left( x \frac{f_1(X)}{f_0(X)} \right) + (1 - \gamma) \rho_{r+1}^n \left( y \frac{f_1(X)}{f_0(X)} \right) \right] \\ &= \gamma E_{\theta_0} \left[ \rho_{r+1}^n \left( x \frac{f_1(X)}{f_0(X)} \right) \right] + (1 - \gamma) E_{\theta_0} \left[ \rho_{r+1}^n \left( y \frac{f_1(X)}{f_0(X)} \right) \right] \\ &= \gamma h_r^n(x) + (1 - \gamma)h_r^n(y). \end{aligned}$$

Por lo tanto,  $h_r^n(z)$  es cóncava para toda  $z \in \mathbb{R}^+$ . Luego,  $\rho_r^n(z)$  también lo es en  $[0, \infty)$ .

Como el límite puntual de una sucesión de funciones cóncavas es cóncava, tenemos que  $\rho_r(z) = \lim_{n \rightarrow \infty} \rho_r^n(z)$  y  $h_r(z) = \lim_{n \rightarrow \infty} h_r^n(z)$  también son cóncavas en  $[0, \infty)$ .

Ahora demostremos la continuidad de todas las funciones.

Como  $[0, \infty)$  es cerrado, entonces por el Teorema 10.1 de [10], sabemos que todas las funciones  $\rho_r^n$ ,  $\rho_r$ ,  $h_r^n$  y  $h_r$  son continuas en el interior. Por lo que falta probar que sean continuas en cero por la derecha. Pero esto se sigue de

$$0 \leq \rho_r(z) \leq \rho_r^n(z) \leq g(z) \rightarrow 0 \quad \text{como } z \rightarrow 0 \text{ para toda } n.$$

Para ver la continuidad en cero para  $h_r^n$  y  $h_r$ , tenemos que

$$\begin{aligned} h_r^n(z) &= E_{\theta_0} \left[ \rho_{r+1}^n \left( z \frac{f_1(X_{r+1})}{f_0(X_{r+1})} \right) \right] \\ &\leq E_{\theta_0} \left[ g \left( z \frac{f_1(X_{r+1})}{f_0(X_{r+1})} \right) \right] \\ &\leq E_{\theta_0} \left[ \lambda_1 z \frac{f_1(X_{r+1})}{f_0(X_{r+1})} \right] \\ &= \lambda_1 z E_{\theta_0} \left[ \frac{f_1(X_{r+1})}{f_0(X_{r+1})} \right] \\ &= \lambda_1 z \sum_{x_{r+1}} \left[ f_0(x_{r+1}) \frac{f_1(x_{r+1})}{f_0(x_{r+1})} \right] \\ &= \lambda_1 z \quad \forall n. \end{aligned}$$

Entonces al hacer  $z \rightarrow 0$  tenemos que  $h_r^n(z) \rightarrow 0$  para toda  $n$ , entonces también  $h_r(0) \rightarrow 0$ . Por lo tanto, queda demostrado este punto.

- ii) Dados  $x, y \in \mathbb{R}^+$  con  $x \leq y$ , queremos ver que se cumple  $\rho_r^n(x) \leq \rho_r^n(y)$ ,  $\rho_r(x) \leq \rho_r(y)$ ,  $h_r^n(x) \leq h_r^n(y)$  y  $h_r(x) \leq h_r(y)$ . Por propiedad de la esperanza es inmediato ver que si  $\rho_r^n(x) \leq \rho_r^n(y)$  y  $\rho_r(x) \leq \rho_r(y)$ , se sigue que  $h_r^n(x) \leq h_r^n(y)$  y  $h_r(x) \leq h_r(y)$ . Por lo que sólo hay que probarlo para  $\rho_r^n$  y  $\rho_r$ .

Para  $r = n$  tenemos

$$\rho_n^n(x) = g(x) = \min\{\lambda_0, \lambda_1 x\} \leq \min\{\lambda_0, \lambda_1 y\} = g(y) = \rho_n^n(y).$$

Supongamos que  $\rho_{r+1}^n$  es no decreciente en  $[0, \infty)$  para algún  $r + 1$ . Entonces,

$$\rho_r^n(x) = \min \left\{ g(x), g(x)p_r + (1 - p_r) \left( c + E_{\theta_0} \left[ \rho_{r+1}^n \left( x \frac{f_1(X)}{f_0(X)} \right) \right] \right) \right\}$$

$$\begin{aligned}
&\leq \min \left\{ g(y), g(y)p_r + (1 - p_r) \left( c + E_{\theta_0} \left[ \rho_{r+1}^n \left( x \frac{f_1(X)}{f_0(X)} \right) \right] \right) \right\} \\
&\leq \min \left\{ g(y), g(y)p_r + (1 - p_r) \left( c + E_{\theta_0} \left[ \rho_{r+1}^n \left( y \frac{f_1(X)}{f_0(X)} \right) \right] \right) \right\} \\
&= \rho_r^n(y).
\end{aligned}$$

Por lo tanto, también se cumple para  $r$ . Sabemos que el límite puntual de funciones no decrecientes es una función no decreciente, así que  $\rho_r(z)$  también lo es. Luego, también la esperanza bajo  $\theta_0$  de las funciones  $\rho_r^n$  y  $\rho_r$  son no decrecientes en  $[0, \infty)$ , es decir,  $h_r^n$  y  $h_r$ .

iii) Probemos primero que  $\rho_r^n(0) = \rho_r(0) = 0$  y  $\rho_r^n(\infty) = \rho_r(\infty) = \lambda_0$ . Entonces para  $r = n$  tenemos que  $\rho_n^n(0) = g(0) = \min\{\lambda_0, 0\} = 0$  y  $\lim_{z \rightarrow \infty} \rho_n^n(z) = \lim_{z \rightarrow \infty} g(z) = \lim_{z \rightarrow \infty} \min\{\lambda_0, \lambda_1 z\} = \lambda_0$ . Supongamos que se cumple en ambos casos para  $r + 1$ . Entonces vemos que

$$\begin{aligned}
\rho_r^n(0) &= \min \left\{ g(0), g(0)p_r + (1 - p_r) \left( c + E_{\theta_0} [\rho_{r+1}^n(0)] \right) \right\} \\
&= \min \{0, c(1 - p_r)\} = 0 \quad \forall n.
\end{aligned}$$

Por lo que también se cumple para  $r$ . Pero además, hemos probado que  $E_{\theta_0}[\rho_{r+1}^n(0)] = h_r^n(0) = 0$  para toda  $1 \leq r \leq n$ .

Por otra parte,

$$\begin{aligned}
\rho_r^n(\infty) &= \lim_{z \rightarrow \infty} \rho_r^n(z) = \lim_{z \rightarrow \infty} \min \left\{ g(z), g(z)p_r + (1 - p_r) \left( c + E_{\theta_0} \left[ \rho_{r+1}^n \left( z \frac{f_1(X)}{f_0(X)} \right) \right] \right) \right\} \\
&= \min \left\{ \lim_{z \rightarrow \infty} g(z), \lim_{z \rightarrow \infty} g(z)p_r + (1 - p_r) \right. \\
&\quad \left. \times \left( c + \lim_{z \rightarrow \infty} E_{\theta_0} \left[ \rho_{r+1}^n \left( z \frac{f_1(X)}{f_0(X)} \right) \right] \right) \right\} \\
&= \min \left\{ \lambda_0, \lambda_0 p_r + (1 - p_r) \left( c + \lim_{z \rightarrow \infty} \sum_x f_{\theta_0}(x) \rho_{r+1}^n \left( z \frac{f_1(X)}{f_0(X)} \right) \right) \right\}.
\end{aligned}$$

Como la suma en la última expresión es finita, entonces se tiene que

$$\lim_{z \rightarrow \infty} \sum_x f_{\theta_0}(x) \rho_{r+1}^n \left( z \frac{f_1(X)}{f_0(X)} \right) = \sum_x f_{\theta_0}(x) \lim_{z \rightarrow \infty} \rho_{r+1}^n \left( z \frac{f_1(X)}{f_0(X)} \right).$$

Luego,

$$\begin{aligned}
 \rho_r^n(\infty) &= \lim_{z \rightarrow \infty} \rho_r^n(z) = \min \left\{ \lambda_0, \lambda_0 p_r + (1 - p_r) \left( c + \sum_x f_{\theta_0}(x) \lambda_0 \right) \right\} \\
 &= \min \{ \lambda_0, \lambda_0 p_r + (1 - p_r)(c + \lambda_0) \} \\
 &= \min \{ \lambda_0, \lambda_0 + c(1 - p_r) \} = \lambda_0 \quad \forall n.
 \end{aligned}$$

Por lo tanto, también se cumple para  $r$ . Pero también hemos probado, de manera indirecta, que

$$\lim_{z \rightarrow \infty} h_r^n(z) = \lim_{z \rightarrow \infty} \sum_x f_{\theta_0}(x) \rho_{r+1}^n \left( z \frac{f_1(X)}{f_0(X)} \right) = \sum_x f_{\theta_0}(x) \lim_{z \rightarrow \infty} \rho_{r+1}^n \left( z \frac{f_1(X)}{f_0(X)} \right) = \lambda_0,$$

para toda  $1 \leq r \leq n$ .

Por ii), sabemos que cualquier  $\rho_r^n$  es no decreciente y acotada superiormente por  $\lambda_0$ , entonces existe  $\lim_{z \rightarrow \infty} \rho_r(z) = \lambda$ . Por lo que si pasamos al límite cuando  $z \rightarrow \infty$  en (3.7), tenemos

$$\begin{aligned}
 \lambda &= \lim_{z \rightarrow \infty} \min \left\{ g(z), g(z) p_r + (1 - p_r) \left( c + E_{\theta_0} \left[ \rho_{r+1} \left( z \frac{f_1(X)}{f_0(X)} \right) \right] \right) \right\}, \\
 &= \min \left\{ \lim_{z \rightarrow \infty} g(z), \lim_{z \rightarrow \infty} g(z) p_r + (1 - p_r) \left( c + \lim_{z \rightarrow \infty} E_{\theta_0} \left[ \rho_{r+1} \left( z \frac{f_1(X)}{f_0(X)} \right) \right] \right) \right\} \\
 &= \min \left\{ \lambda_0, \lambda_0 p_r + (1 - p_r) \left( c + \lim_{z \rightarrow \infty} \sum_x f_{\theta_0}(x) \rho_{r+1} \left( z \frac{f_1(X)}{f_0(X)} \right) \right) \right\} \\
 &= \min \left\{ \lambda_0, \lambda_0 p_r + (1 - p_r) \left( c + \sum_x f_{\theta_0}(x) \lim_{z \rightarrow \infty} \rho_{r+1} \left( z \frac{f_1(X)}{f_0(X)} \right) \right) \right\} \\
 &= \min \{ \lambda_0, \lambda_0 p_r + (1 - p_r)(c + \lambda) \}
 \end{aligned}$$

Sabemos que  $\lambda \leq \lambda_0$ , supongamos que  $\lambda < \lambda_0$ , entonces

$$\lambda = \min \{ \lambda_0, \lambda_0 p_r + (1 - p_r)(c + \lambda) \} = \lambda_0 p_r + (1 - p_r)(c + \lambda).$$

Simplificando llegamos a que

$$(1 - p_r)c + p_r(\lambda_0 - \lambda) = 0.$$

Pero al ser  $c > 0$ ,  $\lambda_0 - \lambda > 0$  y  $p_r \in [0, 1]$ , llegamos que  $p_r = 1$  y  $p_r = 0$ , lo cual es una contradicción.

Por lo tanto,  $\lambda = \lambda_0$ . Esto concluye la demostración.



Una vez que sabemos qué propiedades tienen las funciones  $\rho_r^n$ ,  $\rho_r$ ,  $h_r^n$  y  $h_r$ , queremos ver qué sucede en la etapa  $r$ . Por ejemplo, en esta etapa  $r$ , el paro será dado por la desigualdad

$$g(z) \leq c + h_r(z),$$

evaluada para  $z = Z_r$ .

Por el Lema 3.3, sabemos que la función  $h_r$  es continua, no decreciente y cóncava en  $[0, \infty)$ , pero además para  $h_r$  y  $g$ , tenemos el siguiente.

**Lema 3.4.** *Para toda  $z \in \mathbb{R}^+$  se cumple*

$$h_r(z) \leq g(z).$$

*Demostración.*

Sabemos que  $h_r(z) = E_{\theta_0} \left[ \rho_{r+1} \left( z \frac{f_1(X)}{f_0(X)} \right) \right]$ . Ahora bien, por (3.7) tenemos que

$$\rho_{r+1} \left( z \frac{f_1(X)}{f_0(X)} \right) \leq g \left( z \frac{f_1(X)}{f_0(X)} \right).$$

Entonces, tomando esperanza bajo la hipótesis nula en lo anterior tenemos

$$h_r(z) = E_{\theta_0} \left[ \rho_{r+1} \left( z \frac{f_1(X)}{f_0(X)} \right) \right] \leq E_{\theta_0} \left[ g \left( z \frac{f_1(X)}{f_0(X)} \right) \right].$$

Al ser  $g$  cóncava y como  $E_{\theta_0} \left[ g \left( z \frac{f_1(X)}{f_0(X)} \right) \right] \leq \lambda_1 z < \infty$  para toda  $z, \lambda \in \mathbb{R}^+$ , podemos aplicar la desigualdad de Jensen a la parte derecha de la desigualdad anterior. En consecuencia tenemos

$$E_{\theta_0} \left[ g \left( z \frac{f_1(X)}{f_0(X)} \right) \right] \leq g \left( E_{\theta_0} \left[ z \frac{f_1(X)}{f_0(X)} \right] \right) = g \left( z E_{\theta_0} \left[ \frac{f_1(X)}{f_0(X)} \right] \right) = g(z), \quad \forall z \in \mathbb{R}^+,$$

toda vez que  $E_{\theta_0} \left[ \frac{f_1(X)}{f_0(X)} \right] = \sum_x f_0(x) \frac{f_1(x)}{f_0(x)} = 1$ .



Por el lema anterior sabemos que  $g(z) - h_r(z) \geq 0$  para toda  $z \in \mathbb{R}^+$ . Entonces, para detener el experimento secuencial en la etapa  $r$  debe de suceder que  $0 \leq g(Z_r) - h_r(Z_r) \leq c$ , en caso contrario continuamos con el experimento.

En lo que sigue queremos demostrar que existe un único  $z \in [0, \lambda_0/\lambda_1]$  que es solución de  $g(z) - h_r(z) = c$ , siempre que  $c$  esté en la imagen de  $(g - h_r)I_{\{[0, \lambda_0/\lambda_1]\}}$ . De igual manera, que existe un único  $z \in [\lambda_0/\lambda_1, \infty)$  que es solución de  $g(z) - h_r(z) = c$ , si  $c$  está en la imagen de  $(g - h_r)I_{\{[\lambda_0/\lambda_1, \infty)\}}$ .

La Figura 3.1 muestra la forma que tienen las funciones  $g(z)$  y  $h_r(z)$ .

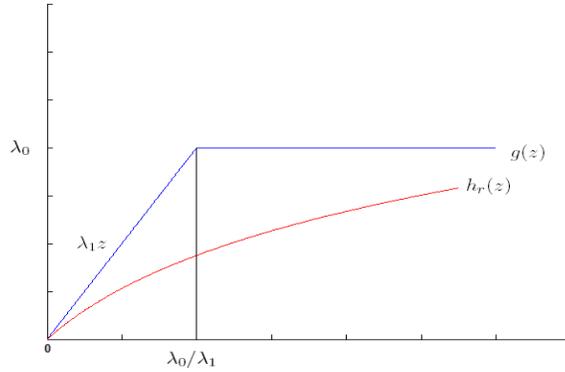


Figura 3.1: Gráfica de  $g(z) = \min\{\lambda_0, \lambda_1 z\}$  y  $h_r(z)$ .

**Lema 3.5.** Sean  $L_r, R_r : [0, \infty) \rightarrow \mathbb{R}^+$  definidas como

$$L_r(z) = \lambda_1 z - h_r(z) \quad (3.11)$$

y

$$R_r(z) = \lambda_0 - h_r(z), \quad \forall z \in \mathbb{R}^+. \quad (3.12)$$

Entonces

- i)  $L_r$  y  $R_r$  son continuas y convexas en  $[0, \infty)$ ,
- ii)  $L_r(0) = R_r(\infty) = 0$ ,  $L_r(\infty) = \infty$ ,  $R_r(0) = \lambda_0$  y  $L_r(\lambda_0/\lambda_1) = R_r(\lambda_0/\lambda_1)$ .
- iii) Si  $x < y$  y  $L_r(x) > 0$ , entonces  $L_r(x) < L_r(y)$ . Por otro lado si  $x < y$  y  $R_r(y) > 0$ , entonces  $R_r(x) > R_r(y)$ .

*Demostración.*

Observemos que  $L_r$  y  $R_r$  son en efecto funciones no negativas. Esto se tiene por el Lema 3.3, ya que  $0 \leq g(z) - h_r(z) \leq \lambda_0 - h_r(z)$  y  $0 \leq g(z) - h_r(z) \leq \lambda_1 z - h_r(z)$  para toda  $z \in [0, \infty)$ .

- i) Por el Lema 3.3 sabemos que  $h_r$  es continua y al ser  $g$  continua, entonces  $L_r$  y  $R_r$  también lo son. Como  $\lambda_1 z$  y  $\lambda_0$  son cóncavas y convexas en  $[0, \infty)$  y por el Lema 3.3  $h_r$  es cóncava, luego  $-h_r$  es convexa. Por lo tanto,  $L_r$  y  $R_r$  son convexas.
- ii) Haciendo uso nuevamente del Lema 3.3 es inmediato ver que  $L_r(0) = \lambda_0 0 - h_r(0) = 0 - 0 = 0$ ,  $R_r(\infty) = \lambda_0 - h_r(\infty) = \lambda_0 - \lambda_0 = 0$ ,  $L_r(\infty) = \lambda_1 \infty - \lambda_0 = \infty$ , finalmente tenemos que  $R_r(0) = \lambda_0 - h_r(0) = \lambda_0$  y  $L_r(\lambda_0/\lambda_1) = \lambda_1(\lambda_0/\lambda_1) - h_r(\lambda_0/\lambda_1) = \lambda_0 - h_r(\lambda_0/\lambda_1) = R_r(\lambda_0/\lambda_1)$ .
- iii) Probemos primero que dados  $x < y$  con  $L_r(x) > 0$ , entonces  $L_r(x) < L_r(y)$ . Supongamos lo contrario, es decir que  $L_r(x) \geq L_r(y)$ . Sabemos que existe un  $\alpha \in (0, 1)$ , tal que  $x = \alpha 0 + (1 - \alpha)y = (1 - \alpha)y$ . Utilizando la convexidad de  $L_r$  tenemos que

$$L_r(y) \leq L_r(x) = L_r((1 - \alpha)y) \leq (1 - \alpha)L_r(y) < L_r(y) \quad \forall \alpha \in (0, 1),$$

lo cual es una contradicción. Por lo tanto  $L_r(x) < L_r(y)$ .

Finalmente probemos que si  $x < y$  y  $R_r(y) > 0$ , entonces  $R_r(x) > R_r(y)$ . Supongamos que  $R_r(x) \leq R_r(y)$ .

Por la continuidad de  $R_r$  y por el hecho que  $R_r(\infty) = 0$ , podemos tomar un  $z > y$ , de tal forma que  $0 < R_r(z) < R_r(y)/2$ . 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  $\alpha$  tal que  $y = \alpha x + (1 - \alpha)z$ .

Ahora, por la convexidad de  $R_r$ , tenemos que

$$\begin{aligned} R_r(y) &= R_r(\alpha x + (1 - \alpha)z) \leq \alpha R_r(x) + (1 - \alpha)R_r(z) \\ &\leq \alpha R_r(y) + (1 - \alpha)\frac{R_r(y)}{2} \\ &= (1 + \alpha)\frac{R_r(y)}{2} < R_r(y) \quad \text{para toda } \alpha \in (0, 1). \end{aligned}$$

Lo cual no puede ser cierto. Por lo tanto  $R_r(x) > R_r(y)$ . ■

Observemos que la función  $L_r$  es inyectiva a partir del primer  $x > 0$ , tal que  $L_r(x) > 0$ , mientras que  $R_r$  lo es en todo su dominio.

Con lo anterior estamos en condiciones de simplificar la regla de paro  $\psi^*$  dada en el Teorema 2.15 por la regla de paro dada en (3.10), tal como se muestra en el siguiente lema.

**Lema 3.6.** *Dada la regla de paro del Teorema 2.15, existen constantes  $A_r, B_r$  ( $0 < A_r < B_r < \infty$ ) tales que  $\psi^*$  es equivalente a*

$$\psi_r^* = I_{\{Z_r \notin (A_r, B_r)\}} \quad r=1,2,3,\dots, \quad (3.13)$$

con  $Z_r = \frac{f_1^r}{f_0^r}$ .

*Demostración.*

Sabemos que la regla de paro del Teorema 2.15 está dada como

$$\psi_r^* = I_{\{m_r \leq cf_0^r + \sum_{x_{r+1}} V_{r+1}\}}, \quad r = 1, 2, 3, \dots$$

Tomando en cuenta (3.5) llegamos a que  $\psi^*$  es equivalente a (3.8) y por (3.9) tenemos que

$$\psi_r^* = I_{\{g(z) \leq c + h_r(z)\}} \quad \text{para } z = Z_r, r = 1, 2, 3, \dots$$

Sabemos que la regla óptima la tendremos cuando aparezca la primera  $r$ , tal que  $g(z) \leq c + h_r(z)$  para  $z = Z_r$ . Entonces si el experimento continúa en la etapa  $r + 1$ , significa que en la etapa anterior se tuvo  $g(z) > c + h_r(z)$  para  $z = Z_r$  o equivalentemente  $c < g(z) - h_r(z)$  para  $z = Z_r$ .

Por otra parte y haciendo uso del Lema 3.5, tenemos que

$$\begin{aligned} g(z) - h_r(z) &= L_r I_{\{[0, \lambda_0/\lambda_1]\}}(z) + R_r I_{\{[\lambda_0/\lambda_1, \infty)\}}(z) \\ &= L_r I_{\{[0, \lambda_0/\lambda_1]\}}(z) + R_r I_{\{[\lambda_0/\lambda_1, \infty)\}}(z) \quad \text{para toda } z \in \mathbb{R}^+. \end{aligned}$$

Por el mismo lema, sabemos que el máximo de  $g(z) - h_r(z)$  se alcanza en  $\lambda_0/\lambda_1$ , ya que  $L_r$  es creciente en  $[0, \lambda_0/\lambda_1]$  y  $R_r$  es decreciente en  $[\lambda_0/\lambda_1, \infty)$ , mas aún,  $L_r$  es estrictamente creciente a partir del primer  $x > 0$  tal que  $L_r(x) > 0$  y  $R_r$  es estrictamente decreciente en todo  $\mathbb{R}^+$  (ver Figura 3.2). Denotemos a  $M = \max_z \{g(z) - h_r(z)\} = L_r(\lambda_0/\lambda_1) = R_r(\lambda_0/\lambda_1)$ . Entonces en la etapa  $r$  se cumple que  $c < M$ .

Haciendo uso nuevamente del Lema 3.5, tenemos que  $L_r$  y  $B_r$  son continuas en  $\mathbb{R}^+$  y al ser  $c > 0$ , podemos aplicar el Teorema del Valor Intermedio, con lo cual sabemos que existe  $x > 0$  tal que  $0 < L_r(x) = c/2$ , de tal manera que  $c \in (c/2, M)$ . Aplicando nuevamente el Teorema del Valor Intermedio, existe  $A_r \in (x, \lambda_0/\lambda_1)$  tal que  $L_r(A_r) = c$  y también existe  $B_r \in (\lambda_0/\lambda_1, \infty)$ , tal que  $R_r(B_r) = c$ . De lo anterior es inmediato ver que se tiene la siguiente relación:  $0 < A_r < B_r < \infty$ .

Por otra parte, al ser  $L_r$  creciente en  $(A_r, \lambda_0/\lambda_1]$ , tenemos que para toda  $z \in (A_r, \lambda_0/\lambda_1)$ ,  $c < g(z) - h_r(z)$  y al ser  $R_r$  decreciente en  $(\lambda_0/\lambda_1, B_r)$ , se cumple que para toda  $z \in (\lambda_0/\lambda_1, B_r)$ ,  $c < g(z) - h_r(z)$ .

Así concluimos que para toda  $z \in (A_r, B_r)$  se cumple que  $c < g(z) - h_r(z)$ .

Por lo tanto, el experimento secuencial terminará si  $z = Z_r \notin (A_r, B_r)$ , es decir, la regla de paro se transforma en (3.13), como se quería probar. ■

Observemos que en el Teorema 2.15, tenemos el caso cuando el horizonte aleatorio  $D$  es acotado. En este caso, sabemos que existe un único  $k \in \mathbb{Z}^+$ , tal que la función de riesgo es uno ( $p_k = 1$ ) y la regla de paro sólo cambia la componente  $k$ -ésima ( $\psi_k^* \equiv 1$ ). Por lo tanto, no importa encontrar las cotas  $A_k, B_k$  del intervalo de continuación  $(A_k, B_k)$ , ya que el experimento secuencial terminará de manera forzada por el horizonte, al ser éste acotado en la etapa  $k$ .

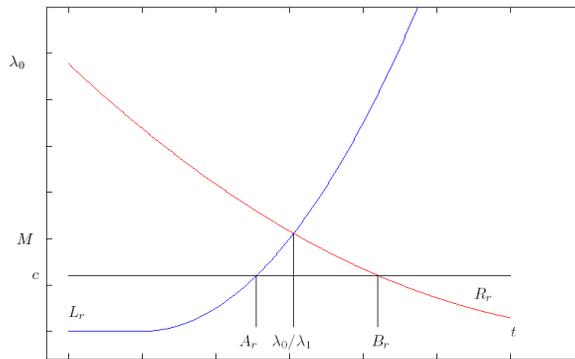


Figura 3.2: Restricción de las funciones  $L_r : [0, \lambda_0/\lambda_1] \rightarrow \mathbb{R}^+$  y  $R_r : (\lambda_0/\lambda_1, \infty) \rightarrow \mathbb{R}^+$  para obtener  $g(z) - h_r(z)$ .

Notemos que en cada etapa  $r$ , se obtienen diferentes cotas de continuación  $A_r$  y  $B_r$  (o también llamado intervalo de continuación  $(A_r, B_r)$ ), ya que dependen del horizonte aleatorio en esa etapa

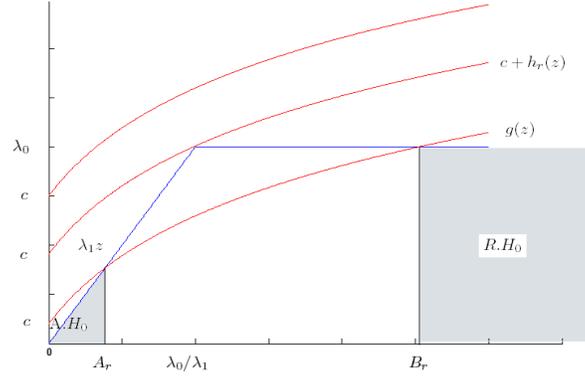


Figura 3.3: Valores  $A_r$  y  $B_r$  para los cuales la regla de paro equivale a  $g(z) \leq c + h_r(z)$  siempre que  $z \notin (A_r, B_r)$ .

$$\left( p_r = \frac{P(D=r)}{P(D \geq r)} \right).$$

Recordemos que por la regla de decisión  $\delta^*$  del Teorema 2.15, aceptaremos  $H_0$  si  $\delta_r^* \equiv 0$ , esto es  $\delta_r^* = I_{\{\lambda_0 f_0^r \leq \lambda_1 f_1^r\}} = 0$ , pero esto ocurre si y sólo si  $\lambda_0 f_0^r > \lambda_1 f_1^r$ , o equivalentemente si  $\lambda_0/\lambda_1 > Z_r$ . Entonces, aceptamos la hipótesis nula  $H_0$  si  $Z_r < \lambda_0/\lambda_1$  y la rechazamos si  $\lambda_0/\lambda_1 \leq Z_r$ .

Podríamos tener el caso que de entrada  $c \geq \lambda_0/\lambda_1$ , entonces  $g(z) \leq c + h_1(z)$  para toda  $z \in \mathbb{R}^+$ , Por lo tanto, no se llevaría a cabo el experimento, pero es un caso que no consideraremos ya que estamos en la clase de pruebas con al menos una observación.

En la Figura 3.3 se muestra cuando se suma la constante de costo unitario  $c > 0$  a la función  $h_r$ . El área de continuación es el intervalo  $(A_r, B_r)$ .

Ahora estamos en condiciones de expresar el Teorema 2.15, pero en términos de razón de probabilidades.

**Corolario 3.7.** Sea  $(\psi^*, \delta^*)$  una prueba secuencial donde  $\delta^* = (\delta_1^*, \delta_2^*, \delta_3^*, \dots)$ , con

$$\delta_r^* = I_{\{\lambda_0/\lambda_1 \leq Z_r\}} = \begin{cases} 1 & \text{si } \lambda_0/\lambda_1 \leq Z_r, r = 1, 2, 3, \dots, \\ 0 & \text{en otro caso,} \end{cases} \quad (3.14)$$

y

$$\psi_r^* = I_{\{Z_r \notin (A_r, B_r)\}}, r = 1, 2, 3, \dots, \quad (3.15)$$

$Z_r = f_1^r/f_0^r$  y  $0 < A_r < B_r < \infty$ . Las constantes  $A_r$  y  $B_r$  cumplen con  $L_r(A_r) = c = R_r(B_r)$  del Lema 3.5. Entonces para cualquier otra prueba secuencial  $(\psi, \delta)$ , tal que

$$\alpha(\psi, \delta) \leq \alpha(\psi^*, \delta^*) \quad y \quad \beta(\psi, \delta) \leq \beta(\psi^*, \delta^*), \quad (3.16)$$

se cumple:

$$E_{\theta_0}[\nu_{\psi^*} \wedge D] \leq E_{\theta_0}[\nu_{\psi} \wedge D]. \quad (3.17)$$

Si  $D$  es un horizonte aleatorio acotado ( $p_k = 1$  para alguna  $k$ ) entonces la regla de paro cambia a  $\psi^* = (\psi_1^*, \psi_2^*, \psi_3^*, \dots, \psi_{k-1}^*, \psi_k^* \equiv 1)$ .

La desigualdad en (3.17) será estricta si alguna de las desigualdades en (3.16) es estricta.

*Demostración.*

Por el Lema 3.6 sabemos existen cotas  $A_r, B_r$  del intervalo de continuación en la etapa  $r$  que cumplen con  $L_r(A_r) = c = R_r(B_r)$  del Lema 3.5, tal que la regla de paro del Teorema 2.15 es equivalente a (3.15). Además, por el mismo lema se tiene que  $0 < A_r < B_r < \infty$ .

Como hemos supuesto que  $f_{\theta_0}^r > 0$  y  $f_{\theta_1}^r > 0$  y sabemos que el caso no trivial es cuando  $\lambda_0 > 0$  y  $\lambda_1 > 0$ . Entonces la regla de decisión es fácil ver que se transforma en

$$\delta_r^* = I_{\{\lambda_0/\lambda_1 \leq Z_r\}}.$$

Así tenemos que las reglas de decisión (3.14) y de paro (3.15) equivalen a las reglas de decisión (2.66) y la regla de paro (2.67) del Teorema 2.15, respectivamente. De esta manera tenemos las mismas condiciones del Teorema 2.15. Por lo tanto, tenemos la misma conclusión que la dada en dicho teorema. ■

Ahora sabemos que existen cotas  $A_r$  y  $B_r$  del intervalo de continuación para el experimento secuencial en la etapa  $r$ , en caso de que el horizonte  $D$  sea no acotado, ya que si el horizonte  $D$  es acotado en la etapa  $k$ , no es necesario hallar dichas cotas (si existen), porque el proceso debe terminar en dicha etapa debido al horizonte y no por la regla en si. Si se llega a parar de manera forzada el experimento debido al horizonte, entonces sólo aplicamos la regla de decisión.

Notemos que si existe  $k$  tal que  $p_k = 1$ , entonces

$$\rho_k(z) = \min \left\{ g(z), g(z)p_k + (1 - p_k)(c + h_k(z)) \right\},$$

cambia a

$$\rho_k(z) = \min \{g(z), g(z)\} = g(z).$$

Lo que significa que  $g(z) \leq c + h_r(z)$  y por lo tanto debe terminar el proceso, como era de esperarse.

## 2

### Caso Particular: Función de Riesgo constante

En esta sección supondremos que la distribución del horizonte aleatorio  $D$  tiene una distribución geométrica, la cual nos llevará a encontrar sólo dos cotas  $A$  y  $B$ , para todo tiempo  $r$ . Lo anterior ocurre debido a que la función de riesgo  $p_k = \frac{P(D=k)}{P(D \geq k)}$ , bajo la distribución geométrica, se vuelve constante, por lo que se simplifican algunos resultados y se parece, en cierta forma, al caso estudiado por Wald [12] con  $p_k = 0$ .

Las pruebas secuenciales en razón de probabilidades con frecuencia se traduce en un ahorro del 50% en el número de observaciones ([12]), por lo que es conveniente su estudio y análisis. En nuestro caso, no es la misma prueba de razón de probabilidades de Wald, debido a que la función de riesgo en nuestro estudio es diferente a cero.

Como mencionamos anteriormente, vamos a suponer que la distribución del horizonte aleatorio tiene una distribución geométrica, es decir,  $P(D = r) = (1 - p)^{r-1}p$ , con  $p \in (0, 1)$ . Sabemos que  $P(D \geq r) = (1 - p)^{r-1}$ , entonces la función de riesgo queda como

$$p_r = \frac{P(D = r)}{P(D \geq r)} = \frac{(1 - p)^{r-1}p}{(1 - p)^{r-1}} = p \quad \text{para toda } r \in \mathbb{N}.$$

Esto nos dice que bajo el supuesto de que el horizonte aleatorio tenga una distribución geométrica, la función de riesgo permanecerá constante en cada etapa  $r$  y de hecho es la única distribución con esta propiedad.

Ahora definamos nuevas funciones que nos darán las funciones  $\rho_r^N$  y su límite que nos ayudarán a simplificar algunos detalles más adelante.

$$\widehat{\rho}_0(z) = g(z), \quad (3.18)$$

donde  $g(z) = \min\{\lambda_0, \lambda_1 z\}$  y recursivamente para  $n=1,2,3,\dots$

$$\widehat{\rho}_n(z) = \min \left\{ g(z), g(z)p + (1-p) \left( c + E_{\theta_0} \left[ \widehat{\rho}_{n-1} \left( z \frac{f_1(X)}{f_0(X)} \right) \right] \right) \right\}. \quad (3.19)$$

De tal forma que si consideramos ahora  $p_r = p$  en (3.3), tenemos que

$$\rho_N^N(z) = \widehat{\rho}_0(z), \rho_{N-1}^N(z) = \widehat{\rho}_1(z), \rho_{N-2}^N(z) = \widehat{\rho}_2(z), \dots, \rho_r^N(z) = \widehat{\rho}_{N-r}(z) \quad \forall z \in \mathbb{R}^+.$$

Esto queda demostrado en el siguiente.

**Lema 3.8.** Sean  $\rho_N^N$  y  $\rho_n^N$  definidas en (3.2) y (3.3), respectivamente. Entonces si  $p_n = p$  en (3.3) para toda  $n = N-1, N-2, \dots, 1$ , se tiene que

$$\rho_r^N(z) = \widehat{\rho}_{N-r}(z) \quad \forall z \in \mathbb{R}^+,$$

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

*Demostración.*

Procedamos por inducción hacia atrás. Sea  $r = N$ , entonces de (3.2) y (3.18), tenemos

$$\rho_N^N(z) = g(z) = \widehat{\rho}_0(z) \quad \text{para toda } z \in \mathbb{R}^+.$$

Veamos que también se cumple para  $r = N-1$ . Sustituyendo a  $p_r = p$  en (3.3), tenemos que

$$\begin{aligned} \rho_{N-1}^N(z) &= \min \left\{ g(z), g(z)p + (1-p) \left( c + E_{\theta_0} \left[ \rho_N^N \left( z \frac{f_1(X)}{f_0(X)} \right) \right] \right) \right\} \\ &= \min \left\{ g(z), g(z)p + (1-p) \left( c + E_{\theta_0} \left[ \widehat{\rho}_0 \left( z \frac{f_1(X)}{f_0(X)} \right) \right] \right) \right\} \\ &= \widehat{\rho}_1(z) \quad \forall z \in \mathbb{R}^+. \end{aligned}$$

Lo último es cierto por (3.19).

Supongamos que para  $r < N$  se cumple que

$$\rho_r^N(z) = \widehat{\rho}_{N-r}(z) \quad \forall z \in \mathbb{R}^+.$$

Probemos que también se cumple para  $r - 1$ .

$$\begin{aligned}
 \rho_{r-1}^N(z) &= \min \left\{ g(z), g(z)p + (1-p) \left( c + E_{\theta_0} \left[ \rho_r^N \left( z \frac{f_1(X)}{f_0(X)} \right) \right] \right) \right\} \\
 &= \min \left\{ g(z), g(z)p + (1-p) \left( c + E_{\theta_0} \left[ \widehat{\rho}_{N-r} \left( z \frac{f_1(X)}{f_0(X)} \right) \right] \right) \right\} \\
 &= \widehat{\rho}_{N-r+1}(z) = \widehat{\rho}_{N-(r-1)}(z) \quad \forall z \in \mathbb{R}^+.
 \end{aligned}$$

Lo cual concluye la demostración. ■

Sabemos que por el Lema 3.2 la sucesión  $\{\rho_r^n\}_n$  es una sucesión de números decrecientes no negativos, por lo que ya sabemos que existe su límite el cual denotamos anteriormente como  $\rho_r(z) = \lim_{n \rightarrow \infty} \rho_r^n(z) \quad \forall z \in \mathbb{R}^+$ . De esta manera tenemos que la sucesión  $\{\widehat{\rho}_{n-r}\}_n$  también tiene límite cuando  $n \rightarrow \infty$ , con  $r \in \mathbb{Z}^+$  fija y  $z \in \mathbb{R}^+$ .

Entonces, por el Lema 3.8, se cumple que

$$\rho_r(z) = \lim_{n \rightarrow \infty} \rho_r^n(z) = \lim_{n \rightarrow \infty} \widehat{\rho}_{n-r}(z) \quad \forall z \in \mathbb{R}^+.$$

Denotemos a este último límite como  $\widehat{\rho}(z)$  para toda  $z \in \mathbb{R}^+$ . Siendo por definición

$$\widehat{\rho}(z) = \lim_{n \rightarrow \infty} \widehat{\rho}_n(z) \quad \forall z \in \mathbb{R}^+. \quad (3.20)$$

Por lo tanto, para toda  $r \in \mathbb{Z}^+$  se cumple la identidad:

$$\rho_r(z) = \widehat{\rho}(z) \quad \forall z \in \mathbb{R}^+. \quad (3.21)$$

Por otra parte, sabemos que  $h_r^n(z) = E_{\theta_0} \left[ \rho_{r+1}^n \left( z \frac{f_1(X)}{f_0(X)} \right) \right]$  para toda  $z \in \mathbb{R}^+$ . Entonces por lo anterior, tenemos que  $h_r^n$  cumple con la siguiente identidad para toda  $n$ :

$$h_r^n(z) = E_{\theta_0} \left[ \widehat{\rho}_{n-(r+1)} \left( z \frac{f_1(X)}{f_0(X)} \right) \right] \quad \forall z \in \mathbb{R}^+$$

Pasando al límite en (3.19), cuando  $n \rightarrow \infty$ , tenemos que para la función  $\widehat{\rho}$  se cumple la siguiente igualdad:

$$\widehat{\rho}(z) = \min \left\{ g(z), g(z)p + (1-p) \left( c + E_{\theta_0} \left[ \widehat{\rho} \left( z \frac{f_1(X)}{f_0(X)} \right) \right] \right) \right\}. \quad (3.22)$$

De la ecuación (3.9), sabemos que  $h_r(z) = E_{\theta_0} \left[ \rho_{r+1} \left( z \frac{f_1(X)}{f_0(X)} \right) \right]$  y por (3.21) tenemos que dada  $r$ ,  $\rho_r(z) = \widehat{\rho}(z)$  para toda  $z \in \mathbb{R}^+$ . Así,

$$h_r(z) = E_{\theta_0} \left[ \widehat{\rho} \left( z \frac{f_1(X)}{f_0(X)} \right) \right] \quad \forall z \in \mathbb{R}^+.$$

Por lo que  $h_r$  no depende de la etapa  $r$ . Entonces si la función de riesgo es constante  $p_r = p$  y abusando de la notación, podemos definir a

$$h(z) = h_r(z) \quad \forall z \in \mathbb{R}^+ \text{ y } r \in \mathbb{Z}^+. \quad (3.23)$$

Con lo anterior tenemos que  $\widehat{\rho}(z)$  de (3.22) también cumple la siguiente identidad:

$$\widehat{\rho}(z) = \min \left\{ g(z), g(z)p + (1-p)(c + h(z)) \right\}.$$

Ahora bien, por el Lema 3.3,  $\widehat{\rho}$  y  $h$  también tienen las propiedades que ahí se enuncian; es decir, que son cóncavas, continuas y no decrecientes en  $[0, \infty)$ . Además,  $\widehat{\rho}(0) = h(0) = 0$  y  $\widehat{\rho}(\infty) = h(\infty) = \lambda_0$ .

Tomando en cuenta (3.23), así como haciendo uso del Lema 3.4, tenemos que

$$h(z) = h_r(z) \leq g(z) \quad \text{para toda } z \in \mathbb{R}^+ \text{ y para toda } r \in \mathbb{Z}^+. \quad (3.24)$$

Veamos el siguiente corolario del Lema 3.6, el cual establece un criterio para obtener cotas fijas  $A, B$  en cada etapa  $r$ .

**Corolario 3.9.** *Dada la regla de paro  $\psi^*$  del Teorema 2.15, entonces existen dos constantes  $A$  y  $B$  con  $0 < A < B < \infty$ , tales que la regla de paro  $\psi^*$  es equivalente a*

$$\psi_r^* = I_{\{Z_r \notin (A, B)\}} \quad r=1,2,3,\dots, \quad (3.25)$$

con  $Z_r = \frac{f_1^r}{f_0^r}$ .

*Demostración.*

Probamos en el Lema 3.5 que la regla de paro  $\psi^*$  es equivalente a

$$\psi_r^* = I_{\left\{ g(Z_r) \leq c + h_r(Z_r) \right\}}$$

cuando la función de riesgo,  $p_r$  es variable. Ahora, si la función de riesgo es constante en toda etapa, entonces  $h_r(z) = h(z)$  para toda  $z \in \mathbb{R}^+$ , como en (3.23). Luego, la regla de paro  $\psi^*$  queda como

$$\psi_r^* = I \left\{ g(Z_r) \leq c + h(Z_r) \right\},$$

con  $Z_r = \frac{f_1^r(x)}{f_0^r(x)}$ .

Por (3.24), sabemos que se cumple la desigualdad  $g(z) - h(z) \geq 0$  para toda  $z \in \mathbb{R}^+$ .

Sean  $L, R : [0, \infty) \rightarrow \mathbb{R}^+$  definidas como

$$L(z) = \lambda_1 z - h(z), \quad \forall z \in \mathbb{R}^+, \quad (3.26)$$

y

$$R(z) = \lambda_0 - h(z), \quad \forall z \in \mathbb{R}^+. \quad (3.27)$$

Nuevamente por (3.23) y dada  $r \in \mathbb{Z}^+$ , tenemos que

$$L(z) = \lambda_1 z - h(z) = \lambda_1 z - h_r(z) = L_r(z), \quad \forall z \in \mathbb{R}^+,$$

y

$$R(z) = \lambda_0 - h(z) = \lambda_0 - h_r(z) = R_r(z), \quad \forall z \in \mathbb{R}^+.$$

De esta manera podemos aplicar el Lema 3.5, con lo que tenemos las siguientes propiedades para  $L$  y  $R$ :

- i)  $L$  y  $R$  son continuas y convexas en  $[0, \infty)$ ,
- ii)  $L(0) = R(\infty) = 0$ ,  $L(\infty) = \infty$ ,  $R(0) = \lambda_0$  y  $L(\lambda_0/\lambda_1) = R(\lambda_0/\lambda_1)$ .
- iii) Si  $x < y$  y  $L(x) > 0$ , entonces  $L(x) < L(y)$ . Por otro lado si  $x < y$  y  $R(y) > 0$ , entonces  $R(x) > R(y)$ .

De hecho sabemos que  $L$  es estrictamente creciente a partir de  $x > 0$ , tal que  $L(x) > 0$ , mientras que  $R$  es estrictamente decreciente en todo su dominio.

Así tenemos que

$$\begin{aligned} g(z) - h(z) &= LI_{\{[0, \lambda_0/\lambda_1]\}}(z) + RI_{\{(\lambda_0/\lambda_1, \infty)\}}(z) \\ &= LI_{\{[0, \lambda_0/\lambda_1]\}}(z) + RI_{\{[\lambda_0/\lambda_1, \infty)\}}(z) \quad \text{para toda } z \in \mathbb{R}^+. \end{aligned}$$

Entonces si estamos en la etapa  $r$  y no paramos, se debe cumplir en ese momento la siguiente desigualdad:

$$c < g(Z_r) - h(Z_r).$$

Sabemos por la propiedad ii) del Lema 3.5 que el máximo de  $g(z) - h(z)$  se alcanza en el valor  $\lambda_0/\lambda_1$  el cual es  $L(\lambda_0/\lambda_1) = R(\lambda_0/\lambda_1) = \lambda_0 - h(\lambda_0/\lambda_1)$ , por lo que se cumplen las desigualdades:

$$c < L(\lambda_0/\lambda_1) \quad \text{y} \quad c < R(\lambda_0/\lambda_1).$$

Aplicando el Teorema del Valor Intermedio, vemos que existen  $A \in (0, \lambda_0/\lambda_1)$  y  $B \in (\lambda_0/\lambda_1, \infty)$  únicos, tales que

$$L(A) = R(B) = c.$$

Vemos que se cumple que  $0 < A < B < \infty$ , así como para toda  $z \in (A, B)$ , se tiene que  $g(z) > c + h(z)$ , porque  $L$  es estrictamente creciente y  $R$  estrictamente decreciente. ■

Con el Corolario 3.9 se establece la existencia de dos constantes  $A, B$  tales que satisfacen la ecuación

$$c + h(z) = g(z) = \min\{\lambda_0, \lambda_1 z\}.$$

El resultado que sigue es un corolario del Teorema 2.15 con  $p_r = p \in (0, 1)$  para toda  $r = 1, 2, 3, \dots$

**Corolario 3.10.** *Sea  $(\psi^*, \delta^*)$  una prueba secuencial donde  $\delta^* = (\delta_1^*, \delta_2^*, \delta_3^*, \dots)$ , con*

$$\delta_r^* = I_{\{\lambda_0/\lambda_1 \leq Z_r\}} = \begin{cases} 1 & \text{si } \lambda_0/\lambda_1 \leq Z_r, \quad r = 1, 2, 3, \dots, \\ 0 & \text{en otro caso,} \end{cases} \quad (3.28)$$

y

$$\psi_r^* = I_{\{Z_r \notin (A, B)\}}, \quad r = 1, 2, 3, \dots, \quad (3.29)$$

$Z_r = f_1^r/f_0^r$  y  $0 < A < B < \infty$ . Las constantes  $A$  y  $B$  cumplen con  $L(A) = c = R(B)$  de (3.26) y (3.27), respectivamente. Entonces para cualquier otra prueba secuencial  $(\psi, \delta)$ , tal que

$$\alpha(\psi, \delta) \leq \alpha(\psi^*, \delta^*) \quad y \quad \beta(\psi, \delta) \leq \beta(\psi^*, \delta^*), \quad (3.30)$$

se cumple:

$$E_{\theta_0}[\nu_{\psi^*} \wedge D] \leq E_{\theta_0}[\nu_{\psi} \wedge D]. \quad (3.31)$$

La desigualdad en (3.31) será estricta si alguna de las desigualdades en (3.30) es estricta.

*Demostración.*

Por el Corolario 3.9 sabemos que si la función de riesgo es constante en toda etapa, entonces para la regla óptima  $\psi^*$  dada en el Teorema 2.15 existen constante  $A, B$  con  $0 < A < B < \infty$  que cumplen con  $L(A) = c = R(B)$  de (3.26) y (3.27), respectivamente. Entonces la regla de paro se convierte en (3.25). La regla de decisión no cambia.

De esta manera tenemos las mismas hipótesis del Teorema 2.15. Por lo tanto, se cumple (3.31). ■

En resumen, el anterior teorema nos dice que al considerar que el horizonte aleatorio viene con una distribución geométrica, sólo existe un intervalo de continuación, el cual es el mismo en cada etapa del experimento.

## 3

### Problema Inverso

En el Capítulo 2 minimizamos la función objetivo de costo  $C(\psi; c) = cE_{\theta_0}[\nu_{\psi} \wedge D]$  encontrando las reglas de decisión y de paro óptimas, considerando un horizonte aleatorio  $D$ , restringido a las condiciones dadas por los errores del tipo I ( $\alpha(\psi, \delta)$ ) y del tipo II ( $\beta(\psi, \delta)$ ) y que incorporamos en la función de Lagrange  $L(\psi, \delta; c, \lambda_0, \lambda_1)$  con el fin de tener la función objetivo de costo sin restricciones:

$$L(\psi, \delta; c, \lambda_0, \lambda_1) = cE_{\theta_0}[\nu_\psi \wedge D] + \lambda_0\alpha(\psi, \delta) + \lambda_1\beta(\psi, \delta).$$

Al minimizar la función de Lagrange, hemos encontrado que la forma de la prueba estadística secuencial óptima está dada por el Teorema 2.15,  $(\psi^*, \delta^*)$ . Es decir,

$$L(\psi^*, \delta^*; c, \lambda_0, \lambda_1) = \min_{(\psi^*, \delta^*)} L(\psi, \delta; c, \lambda_0, \lambda_1),$$

con  $\psi^*$  dada como en (2.67) si para toda  $r \geq 1$  se tiene  $p_r < 1$  y la regla cambia a la forma

$$\psi^* = (\psi_1^*, \psi_2^*, \dots, \psi_{k-1}^*, \psi_k^* \equiv 1),$$

siempre que  $p_k = 1$  para alguna única  $k$  y  $\delta^* = (\delta_1^*, \delta_2^*, \delta_3^*, \dots)$ , con  $\delta_i^*$  como en (2.13).

Ya en este último capítulo analizamos las pruebas secuenciales óptimas en términos de razón de probabilidades, considerando que el horizonte aleatorio  $D$  tiene una distribución cualquiera (lo que significa que la función de riesgo  $p_r$  es variable en cada etapa  $r$ ), convirtiendo la regla de paro (2.67) en la regla dada en (3.8). Así hallamos cotas  $A_r$  y  $B_r$  en cada etapa  $r$ , con lo que la regla de paro se simplificó en la regla dada en (3.15). Luego, al suponer que la distribución del horizonte aleatorio tenía una distribución geométrica con parámetro  $p$ , se simplificó la búsqueda de las cotas aún más, ya que la función de riesgo se volvió constante,  $p_r = p$  en toda etapa  $r$ . Entonces probamos en el Corolario 3.9 que cuando la función de riesgo,  $p_r$  permanece constante en cada etapa, podemos hallar sólo dos constantes  $0 < A < B < \infty$ , las cuales hacen equivalente la regla de paro (3.8) del Teorema 2.15 con la regla de paro (3.25). En otras palabras, dadas las constantes  $c > 0$ ,  $\lambda_0 > 0$  y  $\lambda_1 > 0$ , encontramos dos constantes  $A$  y  $B$  con  $0 < A < B < \infty$ , tales que la regla de paro óptima  $\psi^*$  del Teorema 2.15 es equivalente a la regla de paro dada en (3.29).

Ahora lo que queremos es resolver el problema inverso; es decir, considerando que el horizonte aleatorio tiene una distribución geométrica, o equivalentemente que la función de riesgo es constante y dadas  $A$  y  $B$  con  $0 < A < B < \infty$ , deseamos encontrar constantes  $c, \lambda_0, \lambda_1 \geq 0$  tal que la regla (3.8) sea óptima. Pero es claro que el número de incógnitas a encontrar son excesivas, ya que partimos sólo de dos datos iniciales, por lo que es conveniente fijar el valor de alguna de las variables a encontrar, en nuestro caso será  $\lambda_1$ .

Para este último propósito, pondremos los resultados de la razón de probabilidades que dependerán de  $z \in \mathbb{R}^+$ , así como de  $c, \lambda_0$  y  $\lambda_1 \geq 0$ . Además, supondremos en lo que resta de esta sección final que el horizonte aleatorio  $D$  tiene una distribución geométrica, con lo que la

función de riesgo o tasa de fallos es constante,  $p_r = p$  para toda  $r$  y  $p \in (0, 1)$ . En particular tenemos que

$$\widehat{\rho}(z; c, \lambda_0, \lambda_1) = \min \{g(z; \lambda_0, \lambda_1), g(z; \lambda_0, \lambda_1)p + (1 - p)(c + h(z; c, \lambda_0, \lambda_1))\}, \quad (3.32)$$

donde

$$g(z; \lambda_0, \lambda_1) = \min\{\lambda_0, \lambda_1 z\} \quad \text{y} \quad h(z; c, \lambda_0, \lambda_1) = E_{\theta_0} \left[ \widehat{\rho} \left( z \frac{f_1(X)}{f_0(X)}; c, \lambda_0, \lambda_1 \right) \right], \quad (3.33)$$

siendo  $\widehat{\rho}(z; c, \lambda_0, \lambda_1) = \lim_{n \rightarrow \infty} \widehat{\rho}_n(z; c, \lambda_0, \lambda_1)$ , donde  $\widehat{\rho}_n(z; c, \lambda_0, \lambda_1)$  se define comenzando por

$$\widehat{\rho}_0(z; c, \lambda_0, \lambda_1) = g(z; \lambda_0, \lambda_1) \quad (3.34)$$

y para  $n = 1, 2, \dots$  definimos recursivamente;

$$\widehat{\rho}_n(z; c, \lambda_0, \lambda_1) = \min \{g(z; \lambda_0, \lambda_1), g(z; \lambda_0, \lambda_1)p + (1 - p)(c + h_n(z; c, \lambda_0, \lambda_1))\}, \quad (3.35)$$

donde

$$h_n(z; c, \lambda_0, \lambda_1) = E_{\theta_0} \left[ \widehat{\rho}_{n-1} \left( z \frac{f_1(X)}{f_0(X)}; c, \lambda_0, \lambda_1 \right) \right]. \quad (3.36)$$

Así mismo, para  $z, c, \lambda_0, \lambda_1 \in \mathbb{R}^+$  fijas, tenemos que

$$h(z; c, \lambda_0, \lambda_1) = \lim_{n \rightarrow \infty} h_n(z; c, \lambda_0, \lambda_1).$$

Observemos que de (3.33) y dado  $x \geq 0$ , se cumple la siguiente identidad:

$$g(xz; \lambda_0, \lambda_1) = \min\{\lambda_0, \lambda_1 xz\} = g(x; \lambda_0, \lambda_1 z).$$

Con ayuda de lo anterior, podemos probar el siguiente lema.

**Lema 3.11.** *Dado  $x \geq 0$ , se cumple que*

$$\widehat{\rho}_n(xz; c, \lambda_0, \lambda_1) = \widehat{\rho}_n(x; c, \lambda_0, \lambda_1 z),$$

para toda  $z, c, \lambda_0, \lambda_1 \in \mathbb{R}^+$ .

*Demostración.*

Procedamos la demostración por inducción. Para toda  $x, z, c, \lambda_0, \lambda_1 \in \mathbb{R}^+$ , sabemos que

$$\widehat{\rho}_0(xz; c, \lambda_0, \lambda_1) = g(xz; \lambda_0, \lambda_1) = g(x; \lambda_0, \lambda_1 z) = \widehat{\rho}_0(x; c, \lambda_0, \lambda_1 z).$$

Notemos que

$$h_1(xz; c, \lambda_0, \lambda_1) = E_{\theta_0} \left[ \widehat{\rho}_0 \left( xz \frac{f_1(X)}{f_0(X)}; c, \lambda_0, \lambda_1 \right) \right] = E_{\theta_0} \left[ \widehat{\rho}_0 \left( x \frac{f_1(X)}{f_0(X)}; c, \lambda_0, \lambda_1 z \right) \right] = h_1(x; c, \lambda_0, \lambda_1 z).$$

Entonces para  $n = 1$  tenemos que

$$\begin{aligned} \widehat{\rho}_1(xz; c, \lambda_0, \lambda_1) &= \text{mín} \{g(xz; \lambda_0, \lambda_1), g(xz; \lambda_0, \lambda_1)p + (1-p)(c + h_1(xz; c, \lambda_0, \lambda_1))\} \\ &= \text{mín} \{g(x; \lambda_0, \lambda_1 z), g(x; \lambda_0, \lambda_1 z)p + (1-p)(c + h_1(x; c, \lambda_0, \lambda_1 z))\} \\ &= \widehat{\rho}_1(x; c, \lambda_0, \lambda_1 z). \end{aligned}$$

Supongamos que para  $n$  se cumple que  $\widehat{\rho}_n(xz; c, \lambda_0, \lambda_1) = \widehat{\rho}_n(x; c, \lambda_0, \lambda_1 z)$ .

Probemos que también se cumple para  $n + 1$ .

$$\begin{aligned} \widehat{\rho}_{n+1}(xz; c, \lambda_0, \lambda_1) &= \text{mín} \{g(xz; \lambda_0, \lambda_1), g(xz; \lambda_0, \lambda_1)p + (1-p)(c + h_{n+1}(xz; c, \lambda_0, \lambda_1))\} \\ &= \text{mín} \left\{ g(x; \lambda_0, \lambda_1 z), g(x; \lambda_0, \lambda_1 z)p + (1-p) \right. \\ &\quad \left. \times (c + E_{\theta_0} \left[ \widehat{\rho}_n \left( xz \frac{f_1(X)}{f_0(X)}; c, \lambda_0, \lambda_1 \right) \right]) \right\} \\ &= \text{mín} \left\{ g(x; \lambda_0, \lambda_1 z), g(x; \lambda_0, \lambda_1 z)p + (1-p) \right. \\ &\quad \left. \times (c + E_{\theta_0} \left[ \widehat{\rho}_n \left( x \frac{f_1(X)}{f_0(X)}; c, \lambda_0, \lambda_1 z \right) \right]) \right\} \\ &= \text{mín} \{g(x; \lambda_0, \lambda_1 z), g(x; \lambda_0, \lambda_1 z)p + (1-p)(c + h_{n+1}(x; c, \lambda_0, \lambda_1 z))\} \\ &= \widehat{\rho}_{n+1}(x; c, \lambda_0, \lambda_1 z). \end{aligned}$$

■

Con lo anterior hemos probado, de manera implícita, que  $h_n(xz; c, \lambda_0, \lambda_1) = h_n(x; c, \lambda_0, \lambda_1 z)$  y  $h(xz; c, \lambda_0, \lambda_1) = h(x; c, \lambda_0, \lambda_1 z)$  para toda  $z, x, c, \lambda_0, \lambda_1 \in \mathbb{R}^+$ .

Como mencionamos anteriormente, el número de variables a encontrar en el problema inverso es excesivo, entonces por simplicidad, lo conveniente es suponer que  $\lambda_0 = \lambda > 0, \lambda_1 = 1$  (caso no

trivial). Abusando de la notación y quitando el valor de  $\lambda_1 = 1$  de las ecuaciones (3.32), (3.33), (3.34), (3.35) y (3.36), tenemos

$$\widehat{\rho}(z; c, \lambda) = \min \{g(z; \lambda), g(z; \lambda)p + (1 - p)(c + h(z; c, \lambda))\}, \quad (3.37)$$

$$g(z; \lambda) = \min\{\lambda, z\} \quad \text{y} \quad h(z; c, \lambda) = E_{\theta_0} \left[ \widehat{\rho} \left( z \frac{f_1(X)}{f_0(X)}; c, \lambda \right) \right], \quad (3.38)$$

$$\widehat{\rho}_0(z; c, \lambda) = g(z; \lambda), \quad (3.39)$$

$$\widehat{\rho}_n(z; c, \lambda) = \min \{g(z; \lambda), g(z; \lambda)p + (1 - p)(c + h_n(z; c, \lambda))\} \quad \forall n \geq 1 \quad (3.40)$$

y

$$h_n(z; c, \lambda) = E_{\theta_0} \left[ \widehat{\rho}_{n-1} \left( z \frac{f_1(X)}{f_0(X)}; c, \lambda \right) \right], \quad (3.41)$$

respectivamente.

**Lema 3.12.** *Dadas  $c$  y  $\lambda$  no negativas, se cumple que*

$$c + h(z; c, \lambda) = \inf_{\psi} \left\{ cE_{\theta_0}[\nu_{\psi} \wedge D] + \lambda\alpha(\psi, \delta^*) + z\beta(\psi, \delta^*) \right\}, \quad (3.42)$$

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

*Demostración.*

El Teorema 2.2 nos dice que el problema de minimizar  $L(\psi, \delta; c, \lambda, 1)$  se reduce a encontrar la regla de paro óptima  $\psi^*$  si la regla de decisión está dada por  $\delta_r^* = I_{\{\lambda f_0^r \leq 1 f_1^r\}} = I_{\{\lambda \leq Z_r\}}$  para  $r = 1, 2, 3, \dots$ . Entonces al hacer  $\lambda_0 = \lambda$  y  $\lambda_1 = 1$ , tenemos que

$$\begin{aligned} L(\psi^*; c, \lambda, 1) &= \inf_{\psi} \{L(\psi, \delta^*; \lambda, 1)\} \\ &= \inf_{\psi} \{cE_{\theta_0}[\nu_{\psi} \wedge D] + \lambda\alpha(\psi, \delta^*) + \beta(\psi, \delta^*)\}. \end{aligned} \quad (3.43)$$

Haciendo uso del Teorema 2.12, así como de (3.5), (3.21) y (3.33), tenemos que

$$\begin{aligned} L(\psi^*; c, \lambda, 1) &= c + \sum_{x_1} V_1(x_1) = c + \sum_{x_1} f_0^1 \rho_1(Z_1) \\ &= c + E_{\theta_0}[\rho_1(Z_1)] = c + E_{\theta_0}[\widehat{\rho}(Z_1)] \\ &= c + h(1; c, \lambda, 1) \end{aligned} \quad (3.44)$$

Al igualar las ecuaciones (3.43) y (3.44) tenemos

$$c + h(1; c, \lambda, 1) = \inf_{\psi} \{cE_{\theta_0}[\nu_{\psi} \wedge D] + \lambda\alpha(\psi, \delta^*) + \beta(\psi, \delta^*)\}.$$

Ahora bien, por el Lema 3.11 tenemos que para toda  $x \in \mathbb{R}^+$ , en particular para  $x = 1$ , se cumple que

$$c + h(z; c, \lambda, 1) = c + h(1; c, \lambda, z) = \inf_{\psi} \{cE_{\theta_0}[\nu_{\psi} \wedge D] + \lambda\alpha(\psi, \delta^*) + z\beta(\psi, \delta^*)\}.$$

Finalmente, al considerar la función  $h$  de (3.38), tenemos que  $h(z; c, \lambda) = h(z; c, \lambda, 1)$ , por lo que se cumple (3.42). ■

Notemos que en el lema anterior el ínfimo lo tomamos sobre todas las reglas de paro, donde la regla de decisión  $\delta^*$  es la óptima; es decir, (3.42) la podemos ver como

$$c + h(z; c, \lambda) = \inf_{\psi, \delta} \left\{ cE_{\theta_0}[\nu_{\psi} \wedge D] + \lambda\alpha(\psi, \delta) + z\beta(\psi, \delta) \right\}. \quad (3.45)$$

A partir de (3.42), definamos lo siguiente:

$$h(z; 0, \lambda) = \inf_{\psi} \{ \lambda\alpha(\psi, \delta^*) + z\beta(\psi, \delta^*) \}. \quad (3.46)$$

Cabe mencionar que para el caso truncado también tenemos un resultado parecido al anterior. Haciendo uso del Corolario 2.6, así como de los lemas 3.8 y 3.1 y de la ecuación (3.36), llegamos a que

$$h_N(z; 0, \lambda) = \inf_{\psi} \{ \lambda\alpha(\psi, \delta^*) + z\beta(\psi, \delta^*) \}, \quad (3.47)$$

donde el ínfimo se toma sobre todas las reglas truncadas a nivel  $N$ ; es decir,  $\psi = (\psi_1, \psi_2, \dots, \psi_{N-1}, \psi_N \equiv 1)$ .

**Lema 3.13.** *Sea  $\delta^*$  la regla de decisión del Corolario 3.10. Entonces para toda regla de paro  $\psi$  se cumple que*

$$\inf_{\psi} \{ \lambda\alpha(\psi, \delta^*) + z\beta(\psi, \delta^*) \} = \lambda \sum_{n=1}^{\infty} P_{\theta_0}(Z_n \geq \lambda/z) P(D = n) + z \sum_{n=1}^{\infty} P_{\theta_1}(Z_n < \lambda/z) P(D = n). \quad (3.48)$$

*Demostración.*

Hagamos  $\lambda_0 = \lambda$ ,  $\lambda_1 = z$  y  $Z_n = f_0^n/f_1^n$  en el Corolario 2.13, entonces el resultado es inmediato al hacer uso del Lema 3.12 con  $c = 0$ . ■

**Lema 3.14.** *Sean  $z, \lambda \in \mathbb{R}^+$  fijas. Entonces la función*

$$c + h(z; c, \lambda)$$

*es continua, cóncava y estrictamente creciente en  $c \in [0, \infty)$ .*

*Demostración.*

Por el Lema 3.12, sabemos que  $c + h(z; c, \lambda)$  es el ínfimo de funciones lineales en  $c$  cuando las otras dos variables se mantienen fijas, entonces es cóncava en  $c \in [0, \infty)$ . Luego, por el Teorema 10.1 de [10] es continua en  $(0, \infty)$ . Falta demostrar la continuidad en cero.

Sabemos que

$$\inf_{\psi} \{\lambda\alpha(\psi, \delta^*) + z\beta(\psi, \delta^*)\} \leq \inf_{\psi} \{cE_{\theta_0}[\nu_{\psi} \wedge D] + \lambda\alpha(\psi, \delta^*) + z\beta(\psi, \delta^*)\}.$$

Entonces, para toda regla de paro truncada de la forma  $\psi = (\psi_1 \equiv 0, \psi_1 \equiv 0, \dots, \psi_{N-1} \equiv 0, \psi_N \equiv 1)$  y haciendo uso de (3.48) del Lema 3.13, así como del Corolario 2.7 del Capítulo 2 tenemos que

$$\begin{aligned} 0 &\leq \inf_{\psi} \{cE_{\theta_0}[\nu_{\psi} \wedge D] + \lambda\alpha(\psi, \delta^*) + z\beta(\psi, \delta^*)\} - \inf_{\psi} \{\lambda\alpha(\psi, \delta^*) + z\beta(\psi, \delta^*)\} \\ &\leq cN + \left( \lambda P_{\theta_0}(Z_N \geq \lambda/z) + zP_{\theta_1}(Z_N < \lambda/z) \right) P(D > N) \\ &\quad + \sum_{n=1}^N \left( \lambda P_{\theta_0}(Z_n \geq \lambda/z) + zP_{\theta_1}(Z_n < \lambda/z) \right) P(D = n) \\ &\quad - \lambda \sum_{n=1}^{\infty} P_{\theta_0}(Z_n \geq \lambda/z) P(D = n) - z \sum_{n=1}^{\infty} P_{\theta_1}(Z_n < \lambda/z) P(D = n) \\ &\leq cN + \lambda P_{\theta_0}(Z_N \geq \lambda/z) + zP_{\theta_1}(Z_N < \lambda/z) \\ &\quad - \lambda \sum_{n=N+1}^{\infty} P_{\theta_0}(Z_n \geq \lambda/z) P(D = n) - z \sum_{n=N+1}^{\infty} P_{\theta_1}(Z_n < \lambda/z) P(D = n) \\ &\leq cN + \lambda P_{\theta_0}(Z_N \geq \lambda/z) + zP_{\theta_1}(Z_N < \lambda/z) \end{aligned}$$

Dado  $\epsilon > 0$  tomemos a  $c < \frac{\epsilon}{3N}$ . Como  $\lambda, z \in \mathbb{R}^+$  son fijas podemos usar el Lema 2.14 y tomar a  $N$  suficientemente grande, tal que  $\lambda P_{\theta_0}(Z_N \geq \lambda/z) < \epsilon/3$  y  $zP_{\theta_1}(Z_N < \lambda/z) < \epsilon/3$ . Entonces y en vista de (3.42) y (3.46), tenemos que

$$0 \leq c + h(z; c, \lambda) - h(z; 0, \lambda) < \epsilon.$$

Por lo tanto hemos probado que  $\lim_{c \rightarrow 0} c + h(z; c, \lambda) = h(z; 0, \lambda)$ , así que  $c + h(z; c, \lambda)$  es continua en  $c \in [0, \infty)$ .

Veamos que es estrictamente creciente en  $[0, \infty)$ . Sean  $0 \leq c_1 < c_2$ . Entonces

$$\begin{aligned} c_1 + h(z; c_1, \lambda) &= \inf\{c_1 E_{\theta_0}[\nu_\psi \wedge D] + \lambda\alpha(\psi, \delta^*) + z\beta(\psi, \delta^*)\} \\ &< \inf\{c_2 E_{\theta_0}[\nu_\psi \wedge D] + \lambda\alpha(\psi, \delta^*) + z\beta(\psi, \delta^*)\} \\ &= c_2 + h(z; c_2, \lambda). \end{aligned}$$

■

**Lema 3.15.** Sean  $z, c \in \mathbb{R}^+$  fijas. Entonces la función

$$c + h(z; c, \lambda)$$

es cóncava y creciente en  $\lambda \in [0, \infty)$ , y continua en  $\lambda \in (0, \infty)$

*Demostración.*

Procediendo como en la demostración del Lema 3.14, pero fijando ahora a  $z$  y  $c$ , encontramos que  $c + h(z; c, \lambda)$  cóncava en  $\lambda \in [0, \infty)$  y por Teorema 10.1 de [10] es continua en  $\lambda \in (0, \infty)$ .

Finalmente, sean  $\lambda_1 \leq \lambda_2$ , entonces

$$\begin{aligned} c + h(z; c, \lambda_1) &= \inf\{cE_{\theta_0}[\nu_\psi \wedge D] + \lambda_1\alpha(\psi, \delta^*) + z\beta(\psi, \delta^*)\} \\ &\leq \inf\{cE_{\theta_0}[\nu_\psi \wedge D] + \lambda_2\alpha(\psi, \delta^*) + z\beta(\psi, \delta^*)\} \\ &= c + h(z; c, \lambda_2). \end{aligned}$$

■

**Lema 3.16.** Para toda  $z \in \mathbb{R}^+$  fijo,  $h(z; c, \lambda)$  es continua como función de  $(c, \lambda)$ , para cualquier punto  $(c_0, \lambda_0)$  con  $c_0, \lambda_0 \geq 0$ .

*Demostración.*

Sean  $c_0, \lambda_0 \geq 0$  dos números cualquiera, pero fijos. Entonces para toda  $\lambda_1, \lambda_2$  tales que  $\lambda_1 < \lambda_0 < \lambda_2$  tenemos

$$h(z; c, \lambda_1) \leq h(z; c, \lambda) \leq h(z; c, \lambda_2)$$

si  $\lambda \in [\lambda_1, \lambda_2]$ . Ahora,

$$\lim_{c \rightarrow c_0} h(z; c, \lambda_1) \leq \lim_{c \rightarrow c_0} \inf_{\lambda \rightarrow \lambda_0} h(z; c, \lambda) \leq \lim_{c \rightarrow c_0} \sup_{\lambda \rightarrow \lambda_0} h(z; c, \lambda) \leq \lim_{c \rightarrow c_0} h(z; c, \lambda_2),$$

o equivalentemente

$$h(z; c_0, \lambda_1) \leq \lim_{c \rightarrow c_0} \inf_{\lambda \rightarrow \lambda_0} h(z; c, \lambda) \leq \lim_{c \rightarrow c_0} \sup_{\lambda \rightarrow \lambda_0} h(z; c, \lambda) \leq h(z; c_0, \lambda_2),$$

Ahora, haciendo  $\lambda_1 \rightarrow \lambda_0$  y  $\lambda_2 \rightarrow \lambda_0$  en la anterior desigualdad y notando que

$$\lim_{\lambda \rightarrow \lambda_0} h(z; c_0, \lambda) = h(z; c_0, \lambda_0),$$

obtenemos

$$\lim_{c \rightarrow c_0, \lambda \rightarrow \lambda_0} h(z; c, \lambda) = h(z; c_0, \lambda_0).$$

■

Ahora ya estamos en condiciones de probar el resultado principal de esta última sección.

**Teorema 3.17.** Sean  $A, B$  números reales de tal manera que  $0 < A < B < \infty$ . Entonces, existe  $C \in [A, B]$  tal que la prueba secuencial  $(\psi^{**}, \delta^{**})$ , con

$$\psi_r^{**} = I_{\{Z_r \notin (A, B)\}} \quad (3.49)$$

y

$$\delta_r^{**} = I_{\{C \leq Z_r\}} \quad r=1, 2, 3, \dots, \quad (3.50)$$

es óptima en el siguiente sentido: para cualquier otra prueba estadística secuencial  $(\psi, \delta)$ , tal que

$$\alpha(\psi, \delta) \leq \alpha(\psi^{**}, \delta^{**}) \quad y \quad \beta(\psi, \delta) \leq \beta(\psi^{**}, \delta^{**}) \quad (3.51)$$

se cumple que

$$E_{\theta_0}[\nu_{\psi^{**}} \wedge D] \leq E_{\theta_0}[\nu_{\psi} \wedge D]. \quad (3.52)$$

La desigualdad en (3.52) es estricta si alguna de las desigualdades en (3.51) es estricta.

*Demostración.*

La idea es ver que la prueba  $(\psi^{**} = (\psi_1^{**}, \psi_2^{**}, \psi_3^{**}, \dots), \delta^{**} = (\delta_1^{**}, \delta_2^{**}, \delta_3^{**}, \dots))$  con  $\psi_r^{**}$  y  $\delta_r^{**}$  como en (3.49) y (3.50), respectivamente, es en realidad la prueba  $(\psi^*, \delta^*)$  del Corolario 3.10 con  $\delta_r^* = I_{\{\lambda_0/\lambda_1 \leq Z_r\}} = I_{\{\lambda \leq Z_r\}}$ , tales que dados  $0 < A < B < \infty$  existen  $c \geq 0$  y  $\lambda > 0$  que son raíces de  $c + h(z; c, \lambda) = g(z; \lambda)$  que hacen de la prueba la óptima. Es claro que  $\psi^{**} = \psi^*$ , falta ver que  $\delta^{**} = \delta^*$ .

Entonces, para cualesquiera  $A$  y  $B$  con  $0 < A < B < \infty$ , queremos encontrar  $c$  y  $\lambda \in (A, B)$  tales que

$$c + h(A; c, \lambda) = A \quad (3.53)$$

y

$$c + h(B; c, \lambda) = \lambda. \quad (3.54)$$

Ahora bien, por el Lema 3.16, sabemos que la función  $c + h(z; c, \lambda)$  está definida para toda  $c \geq 0$  y  $\lambda > 0$ , pero definámosla solo para  $\lambda \in [A, B]$  a  $c = c(\lambda)$  como una solución de la ecuación (3.53). Para asegurar que existe  $\lambda \in [A, B]$ , tal que sea solución de (3.53), verifiquemos primero que se cumple la desigualdad:  $h(A; 0, \lambda) \leq A$ , porque la parte izquierda de (3.53) es continua y estrictamente creciente como función de  $c$ . Además,  $c(\lambda)$  es una función continua de  $\lambda$ , como una función implícita (3.53) definida por una función la cual es continua en las dos variables (Lema 3.16).

Si tomamos una prueba en particular  $(\psi', \delta')$  con al menos una muestra, supongamos por ejemplo que  $\psi' = (\psi'_1 \equiv 1)$  y  $\delta' = (\delta'_1 \equiv 0)$ , entonces paramos en la primer observación y aceptamos  $H_0$ , lo cual implica que  $\alpha(\psi', \delta') = P_{\theta_0}(\text{rechazar } H_0 | H_0) = 0$  y  $\alpha(\psi', \delta') = P_{\theta_1}(\text{aceptar } H_0 | H_1) = 1$  y como el ínfimo se toma sobre todas las reglas de paro, entonces por el Lema 3.12 tenemos que

$$h(A; 0, \lambda) = \inf_{\psi, \delta} \left\{ \lambda \alpha(\psi, \delta) + A \beta(\psi, \delta) \right\} \leq \lambda \alpha(\psi', \delta') + A \beta(\psi', \delta') = A.$$

Ahora para probar (3.54), definamos a

$$f(\lambda) = \lambda - h(B; c(\lambda), \lambda) - c(\lambda).$$

Así vemos que  $f$  es una función continua en  $[A, B]$  por ser la composición de funciones continuas.

Probemos ahora que

$$f(A) \leq 0 \quad \text{y} \quad f(B) \geq 0. \quad (3.55)$$

En efecto, tenemos

$$f(A) = A - h(B; c(A), A) - c(A) \leq A - h(A; c(A), A) - c(A) = 0.$$

por (3.53).

Probemos ahora que

$$f(B) = B - h(B; c(B), B) - c(B) \geq 0. \quad (3.56)$$

Tomando en cuenta (3.53), tenemos que para probar (3.56) es equivalente a verificar que se cumple la desigualdad:

$$B - h(B; c(B), B) \geq A - h(A; c(B), B).$$

Pero lo anterior es cierto, ya que por (3.11) del Lema 3.5 tenemos que  $z - h(z; c(B), B)$  es no decreciente para toda  $z \in \mathbb{R}^+$ . Entonces por el Teorema del Valor Intermedio, sabemos que existe  $\lambda \in [A, B]$  tal que  $f(\lambda) = 0$ . Pero también encontramos  $c(\lambda)$  tal que cumple con (3.53), lo cual es equivalente a  $c + h(z; c, \lambda) = g(z; \lambda) = \min\{\lambda, z\}$ .

Si hacemos  $C = \lambda$  en la regla de decisión del Corolario 3.10, entonces  $\delta_r^* = I_{\{C \leq Z_r\}} = \delta_r^{**}$ , por lo que tendremos las mismas condiciones que ahí se enuncian. Por lo tanto, si  $(\psi, \delta)$  es una prueba cualquiera que satisface (3.51), entonces la prueba  $(\psi^{**}, \delta^{**})$  es óptima y cumple con (3.52). Además, la desigualdad (3.52) es estricta si alguna de las desigualdades en (3.51) es estricta. ■

# 4

## Conclusiones

Encontramos la forma de las pruebas secuenciales óptimas para experimentos con horizonte aleatorio, pero la diferencia con el caso clásico fue la regla de paro (horizonte infinito), como era de esperarse, ya que en el tiempo de paro se involucra el horizonte aleatorio. La regla de decisión no sufrió modificación alguna. Al poner las pruebas secuenciales en términos de razón de probabilidades, obtuvimos diferentes intervalos de continuación en cada etapa debido al horizonte aleatorio, en contraste con la teoría clásica donde sólo hay un intervalo de continuación en cada etapa.

Un caso particular se da al suponer que el horizonte aleatorio tiene una distribución geométrica, ya que sólo hay un intervalo de continuación y la regla de paro es igual a la regla de paro de la SPRT, sin embargo, la regla de decisión sí cambia porque el horizonte puede truncar el experimento secuencial y hay que tomar una decisión si la razón está en el intervalo de continuación.

Esta investigación realizada es original y en la literatura sobre análisis secuencial no existen resultados en torno a modelos con horizonte aleatorio con este enfoque. El caso clásico de horizonte infinito (y también finito, siendo éste fijo) resultan casos particulares de la teoría desarrollada. En particular, se tiene una demostración alternativa de la optimalidad de la prueba SPRT de Wald.

Aunque el trabajo es para un caso particular de observaciones discretas con número finito de valores, se puede ver como pauta para el caso general, incluso para observaciones continuas. Se ve la posibilidad de caracterización completa de la estructura de pruebas secuenciales óptimas. También se abre la posibilidad de aplicarla para procesos estocásticos a tiempo discreto.





- [1] D. BHATTACHARJEE AND N. MUKHOPADHYAY *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, 2012.
- [2] H. CRUZ SUÁREZ, R. ILHUICATZI ROLDÁN AND R. MONTES DE OCA. *Markov Decision Processes on Borel Spaces with Total Cost and Random Horizon*. Journal of Optimization Theory and Applications, DOI: 10.1007/s10957-012-0262-8, 2013.
- [3] S. DAYANIK AND A. J. YU. *Reward-Rate Maximization in Sequential Identification under a Stochastic Deadline*. SIAM Journal on Control and Optimization. 2012.
- [4] B. EISENBERG, B. K. GHOSH AND G. SIMONS. *Properties of Generalized Sequential Probability Ratio Tests*. The Annals of Statistics. Vol. 4, No. 2, 237-251, 1976.
- [5] P. I. FRAZIER AND A. J. YU. *Sequential Hypothesis Testing under Stochastic Deadlines*. Advances in Neural Information Processing Systems, Mit Press, Cambridge, MA. 20: 465-72, 2008.
- [6] T. IIDA AND M. MORI. *Markov Decision Processes with Random Horizon*. Journal of the Operations Research Society of Japan. Vol. 39, No. 4, 592-603, 1996.
- [7] G. LORDEN. *Structure of Sequential Tests Minimizing an Expected Sample Size*. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, Probability Theory and Related Fields, Springer Berlin / Heidelberg. Vol. 51, No. 3, 291-302, 1980.
- [8] A. M. MOOD. *Introduction to the Theory of Statistics*. McGraw-Hill, Inc., USA. (1950).

- [9] A. NOVIKOV. *Optimal Sequential Tests for Two Simple Hypotheses Based on Independent Observations*. International Journal of Pure and Applied Mathematics. Vol. 45, No. 2, 291-314, 2008.
- [10] R. T. ROCKAFELLAR. *Convex Analysis*. Princeton University Press, New Jersey, 1970.
- [11] E. SAMUEL-CAHN. *Optimal Stopping With Random Horizon With Application to the Full-Information Best-Choice Problem With Random Freeze*. Journal of the American Statistical Association. Vol. 91 No. 433, 357-364, 1996.
- [12] A. WALD. *Sequential Analysis*. Wiley, New York. 1947.
- [13] A. WALD AND J. WOLFOWITZ. *Optimum Character of the Sequential Probability Ratio Test*. The Annals of Mathematics Statistics. Vol. 19, No. 3, 326-339, 1948.

Universidad Autónoma Metropolitana - Iztapalapa  
División de Ciencias Básicas e Ingeniería

---

---



OPTIMALIDAD DE PRUEBAS SECUENCIALES  
PARA EXPERIMENTOS CON  
HORIZONTE ALEATORIO

TESIS QUE PRESENTA

JUAN LUIS PALACIOS SOTO

PARA OBTENER EL GRADO DE  
MAESTRO EN CIENCIAS EN MATEMÁTICAS  
APLICADAS E INDUSTRIALES

ASESOR: DR. ANDREY NOVIKOV

JURADO CALIFICADOR:

PRESIDENTE: DR. JUAN GONZÁLEZ HERNÁNDEZ

SECRETARIO: DR. RAÚL MONTES DE OCA MACHORRO

VOCAL: DR. ANDREY NOVIKOV

*Juan González Hernández*  
*Raúl Montes de Oca Machorro*  
*Andrey Novikov*

MÉXICO, D.F. MARZO 2014