UNIVERSIDAD AUTÓNOMA METROPOLITANA

Unidad Iztapalapa

División de Ciencias Básicas e Ingeniería

Optimalidad de pruebas de hipótesis secuenciales con grupos de tamaño aleatorio

Tesis que presenta

XÓCHITL ITXEL POPOCA JIMÉNEZ

Para obtener el grado de

Maestra en Ciencias

Matemáticas Aplicadas e Industriales

Asesor:

Dr. Andrey Novikov Новиков

Jurado calificador:

Presidente:	Dr. Juan González Hernández.	IIMAS
Secretario:	Dr. Andrey Novikov.	UAM-I
Vocal:	Dra. Blanca Rosa Pérez Salvador.	UAM-I

México D. F. **Julio de 2012**

Dedicatoria

Hace tres años emprendí un camino con el espíritu fragmentado,
porque dejaste en mí una discontinuidad esencial
con un vacío infinito.
Pero me enseñaste a caminar no importa el sendero
y a tener brío no importa la ventisca.

Ahora el recorrido he culminado,
y he logrado entereza debido a tí.

Te amaré siempre.

A la memoria de mi madre
Ma. Luisa Jiménez Glz. (R.I.P.)

Agradecimientos

Al Dr. *Andrey Novikov* por compartirme su: conocimiento, tiempo y confianza.

A mi padre *L. Daniel Popoca* por su dilección y ser mi base. A mi hermano y hermanas.

A mi esposo *Vladímir Camacho* por todo su apoyo, comprensión y cariño.

A mis compañeros y amigos que me ayudaron y compartieron su tiempo. Aldo, Aracelí, Iván, Jessica, Leonel, Naturaleza y Román.

A todos GRACIAS.

ÍNDICE GENERAL

CAPÍTULO 1

ANÁLISIS SECUENCIAL CON GRUPOS ALEATORIOS

1.1. Introducción

En contraste con los métodos clásicos de la estadística matemática, en los cuales el número de observaciones en la muestra está fijo; el *análisis secuencial* (un campo también de la estadística), trabaja, dado un experimento estadístico, con métodos que permiten la recolección de datos por etapas, típicamente de uno en uno. Otra característica de estos métodos estriba en el hecho de que el tiempo del experimento para el cual se determina el número de observaciones (*tiempo de paro*) es aleatorio.

La teoría moderna del análisis secuencial surgió en respuesta a las demandas de toma de muestras más eficientes en procesos de inspección durante la segunda guerra mundial.

Los primeros desarrollos del nuevo campo fueron introducidos por Abraham Wald (1902-1950) durante la década de los cuarenta del siglo pasado. Wald mostró que el utilizar métodos secuenciales en problemas de pruebas de dos hipótesis simples (con observaciones independientes) conlleva a un número de observaciones más pequeño, en promedio, que aquellos métodos que utilizan una muestra de tamaño fijo [18].

Uno de los tópicos fundamentales desarrollados por Wald en este campo es la *prueba secuencial de razón de probabilidades* (*SPRT*, por sus siglas en inglés, *Sequential Probability Ratio Test*) [19]. Dicha prueba es similar a una prueba de hipótesis basada en una muestra de tamaño fijo (Neyman-Pearson) en la estadística inferencial "clásica", con la diferencia de que ésta se aplica a una colección de datos (muestra) dados de cierto tamaño y la primera se aplica a datos tomados, en principio, de uno por uno, esto es, en orden secuencial.

Una propiedad para la *SPRT* demostrada por Wald y Wolfowitz es que dicha prueba es óptima, en el sentido de que, bajo ciertas condiciones, ella conlleva en un experimento a un tiempo de paro mínimo y, también, a un número de observaciones menor (véase [20]). Este resultado es conocido como *la optimalidad de la SPRT*. Una vez terminado el experimento estadístico secuencial, se tiene que hacer la conclusión al igual que en la estadística inferencial en los mismos términos y con las mismas interpretaciones.

Entonces, una ventaja de utilizar procedimientos secuenciales es que, en promedio, estos requieren de un número menor de observaciones con la misma confiabilidad que en el procedimiento no secuencial. Lo cual conlleva a aplicaciones con menores costos y/o tiempos. Entre algunas de las aplicaciones se encuentran: análisis clínicos [21], control de calidad, técnicas de confiabilidad, economía, finanzas, psicología, entre otras.

Una condición para la optimalidad de la SPRT es que los costos de las etapas del experimento sean proporcionales al número de observaciones en ellas. Mas, existen aplicaciones (*e. g.* ensayos clínicos) en las cuales el costo de una etapa con n unidades experimentales no es proporcional al número de observaciones ($c \cdot n$, donde c es el costo unitario), sino, por ejemplo, es de la forma $c_0 + nc_1$. En casos como éste la SPRT pierde su optimalidad (véase [4]).

Otra condición para la optimalidad de la prueba estadística secuencial es que cada etapa tenga una sola observación [13]. Sin embargo, también es natural considerar planes de muestreo en los que, en cada etapa, en lugar de una sola observación se requiere de un grupo de observaciones, estos planes secuenciales son más generales y tienen la característica de que las observaciones se toman en grupos de tamaño fijo (*group-sequential test*) [5], dichos planes se usaban para atacar problemas de costos no lineales.

Otro enfoque más avanzado sobre costos no proporcionales se puede ver en los trabajos de Schmitz [15] y Cressie y Morgan [4] en 1993, en estas investigaciones se describen procedimientos más generales que los del análisis secuencial clásico: *procedimientos secuencialmente planeados*, dichos procesos consideran grupos de observaciones cuyo tamaño es aleatorio, sin embargo el tamaño de cada grupo depende de las observaciones tomadas en etapas previas. Estos procedimientos resultan más eficientes que los métodos clásicos (véase [4]), pero la estructura de los procedimientos óptimos resulta compleja, y en muchos casos no se pueden determinar de manera explícita (veáse [15]).

En el 2008 Nitis Mukhopadhyay y Basil de Silva [10] presentaron una investigación práctica sobre una “generalización” de la SPRT a la cual llamaron *prueba secuencial aleatoria de la razón de probabilidades (RSPRT: Random Sequential Probability Ratio*

Test). Dicha prueba fue aplicada a experimentos secuenciales para los cuales: el tamaño de los grupos de observaciones no tiene, necesariamente, una medida fija, sino que ésta es aleatoria; y la distribución de probabilidad de dichos tamaños no es fija. El objetivo, en dicha investigación, consiste en decidir entre dos hipótesis simples (la nula y la alternativa) con sus respectivas preasignada probabilidades de error tipo I y tipo II . Sin embargo, en este trabajo se dejaron abiertas algunas preguntas sobre las propiedades de *optimalidad* de la nueva RSPRT, esto se observa en la siguiente cita:

> *3.4 Preguntas abiertas acerca de la optimalidad de la RSPRT.*
>
> *. . . ha estado en nuestras mentes, pero aún no estamos en la posición de observar con más profundidad nuestro propósito sobre la RSPRT [. . .], esperamos examinar algunas de las propiedades de optimalidad, monotonía y eficiencia, para la RSPRT.*[1]
>
> *3.4 Open questions regarding optimality of RSPRT*
> *. . . has been on our minds, but we have not yet positioned ourselves to look deeper into our proposed RSPRT [. . .], we hope to examine some of the optimality, monotonicity, and efficiency properties, if any, for the RSPRT.*
>
> [Íbid, p. 437]

Este nuevo enfoque en el cual las observaciones están en grupos cuyo tamaño es aleatorio y la distribución de estos no es fija, lo podemos llamar como *planes de muestreo aleatorizados* y se puede decir que estos son una extensión de las pruebas secuenciales por grupos [2].

[1] Traducción del autor.
[2] Group-sequential test.

Nuestro trabajo tiene como objetivo principal investigar *los procedimientos óptimos de pruebas* cuando las observaciones vienen en grupos de tamaño aleatorio. Esto es, en el contexto de un experimento estadístico secuencial para problemas de pruebas de dos hipótesis simples, donde el tamaño de los grupos para la siguiente etapa es aleatorio y su distribución no depende de las observaciones recolectadas hasta esta etapa; nos interesa investigar la estructura del procedimiento *óptimo* en la prueba. Para evitar complicaciones técnicas, sólo se verá el caso de observaciones discretas con un número finito de valores (como, por ejemplo, observaciones tipo Bernoulli). La metodología propuesta es la desarrollada en el trabajo de Novikov (veáse [13]) (una versión más avanzada de la misma esta publicada en [11]).

Entonces en nuestra investigación se trabajará este nuevo enfoque y se buscará la estructura del procedimiento óptimo bajo las siguientes condiciones:

- Las observaciones del experimento tienen un número finito de valores;
- las observaciones vienen en grupos cuyo tamaño es aleatorio;
- la distribución de los tamaños de los grupos es fija; y
- el costo de los grupos, en cada etapa, tiene una estructura general.

Investigaciones previas muestran que el enfoque sobre la optimalidad de la prueba secuencial con grupos de tamaño aleatorio no está desarrollada, por tal nuestra investigación es "pionera" en este sentido. Cabe mencionar que el procedimiento óptimo bajo el contexto mencionado, conlleva el obtener una base teórica para atacar problemas como los planteados en investigaciones anteriores (*e. g.* Cressie y Morgan [4]); porque nuestro resultado provee un plan de muestreo aleatorizado en concreto para ser aplicado en dichos problemas. En particular, el resultado nos permitiría demostrar la optimalidad de la prueba RSPRT planteado por Mukhopadhyay *et al* [10], salvo hipótesis diferentes a las marcadas en el artículo.

1.2. Experimento secuencial con grupos de tamaño aleatorio

Piense el lector que encontramos en un experimento estadístico con un problema de prueba de hipótesis simples de la forma: $H_0 : \theta = \theta_0 \ v.\ s.\ H_1 : \theta = \theta_1$ con $\theta_0 \neq \theta_1$. Nos interesan los procedimientos de prueba llamados *secuenciales*; esto significa que el número de observaciones, grupos para nuestro caso, que se van a analizar no queda definido con anticipación, sino que se determina en el curso del experimento, dependiendo de las observaciones obtenidas en cada etapa y donde el tamaño del grupo es aleatorio. En el caso en que todos los "grupos" son de tamaño uno estamos en el contexto clásico del análisis secuencial. Cuando los grupos tienen un número de tamaño fijo (por ejemplo: grupos de dos observaciones, de cinco observaciones, etc.), nos encontramos en el caso conocido como análisis secuencial por grupos (*group-sequential*). Se considerará un contexto general con un *experimento secuencial con grupos de tamaño aleatorio.*

En este contexto, se hace un análisis de los grupos de observaciones de tamaño variable, y nos auxiliaremos de dos reglas que nos ayudan a detener el experimento y a tomar una decisión final con respecto a nuestras hipótesis, estas reglas son conocidas respectivamente como: *regla de paro* y *regla de decisión* (terminal).

Un experimento secuencial se representa por un conjunto de variables aleatorias independientes e idénticamente distribuidas $X_1, X_2, \ldots, X_k, \ldots$ que toman valores en un conjunto finito $\mathcal{X}$, con su función de probabilidad dada por $f_\theta(x) = P_\theta(X = x)$ [3] agrupados secuencialmente en conjuntos de tamaño variable. El proceso de conteo que genera el tamaño de los conjuntos secuenciales y el proceso que genera las variables aleatorias X_i son independientes.

Supongamos que los tamaños de los grupos consecutivos $\nu_1, \nu_2, \ldots, \nu_i, \ldots$ son variables aleatorias independientes e idénticamente distribuidas, cuya función de probabilidad está dada por $P(\nu_i = k) = q_k \geq 0$ con $i = 1, 2, \ldots$ y $k = 0, 1, 2, \ldots$. Nótese que se están considerando grupos de tamaño cero, ya que es posible que un grupo no tenga observaciones. Suponemos, también, que los tamaños de los grupos no exceden un número dado; para lo cual sea $\mathcal{G}$ un subconjunto finito cualquiera $\mathcal{G} \subset \{0, 1, 2, \ldots\}$, por tanto se tiene que: $\sum\limits_{k \in \mathcal{G}} q_k = 1$. Con base en las notaciones anteriores se tendrá un experimento estadístico secuencial en el cual se tienen grupos de observaciones con determinados tamaños aleatorios, denotados por $X_1^{(\nu_1)}, X_2^{(\nu_2)}, \ldots, X_k^{(\nu_k)}, \ldots$.

De esta manera, si $n_1, n_2, \ldots$ es una realización de las variables aleatorias $\nu_1, \nu_2, \ldots$; entonces para cada uno de estos números $(n_i \in \mathcal{G}, i \geq 1)$ se tiene un vector de datos de la forma: $x_1^{(n_1)} = (x_{11}, x_{12}, \ldots, x_{1n_1})$, $x_2^{(n_2)} = (x_{21}, x_{22}, \ldots, x_{2n_2})$, etcétera, donde cada $x_i^{(n_i)} \in \mathcal{X}^{n_i}$. Por definición, el grupo sin observaciones, de denota por $x^{(0)} = ()$ ("el vector sin componentes"). Ahora sea el vector $\boldsymbol{n} = (n_1, n_2, \ldots, n_k) \in \mathcal{G}^k$ una realización del vector de variables aleatorias $\boldsymbol{\nu} = (\nu_1, \nu_2, \ldots, \nu_k)$, para cualquier entero $k \geq 1$, donde las componentes de $\boldsymbol{n}$ son los respectivos tamaños de los grupos consecutivos. Con base en esto se define el vector $(\boldsymbol{n} : s)$ como el subvector de los primeros s elementos del vector $\boldsymbol{n}$ como: $(\boldsymbol{n} : s) = (n_1, n_2, \ldots, n_s)$. Definamos ahora un concepto importante para nuestro trabajo, a saber, *regla de paro.*

[3]Donde $\sum\limits_{x \in \mathcal{X}} f_\theta(x) = 1$.

1.3. Reglas de paro

Una regla de paro es una familia de funciones indicadoras, donde cada componente de la familia toma sólo uno de dos valores, cero 0 o 1; la definición formal es:

Definición 1.

Una *regla de paro* ψ es una familia de funciones indicadoras:

$$\left\{\psi_{\boldsymbol{n}} : \mathcal{X}^{n_1} \times \mathcal{X}^{n_2} \times \cdots \times \mathcal{X}^{n_k} \to \{0,1\},\ \text{con } \boldsymbol{n} \in \mathcal{G}^k \text{ y } k \geq 1\right\}$$

Donde las componentes de la regla de paro tienen la siguiente aplicación:

$$\text{si } \boldsymbol{n} \in \mathcal{G}^k, \quad \psi_{\boldsymbol{n}} = \psi_{\boldsymbol{n}}\left(x_1^{(n_1)}, x_2^{(n_2)}, \ldots, x_k^{(n_k)}\right).$$

Y la interpretación de la regla de paro es la siguiente:

Primero se observa el primer grupo de n_1 observaciones, obteniendo los datos $(x_{11}, \ldots, x_{1n_1}) = x_1^{(n_1)}$. A estos datos se les aplica la función $\psi_{\boldsymbol{n}}$ donde $\boldsymbol{n} = (n_1) \in \mathcal{G}$. Si $\psi_{\boldsymbol{n}}(x_1^{(n_1)}) = 1$ entonces el experimento termina en esta primera etapa y la decisión de rechazar o aceptar H_0 se toma exclusivamente con base en $\left(n_1, x_1^{(n_1)}\right)$. En caso contrario $\left(\textit{es decir}, \text{ si } \psi_{\boldsymbol{n}}(x_1^{(n_1)}) = 0\right)$, el experimento continua, pasando a la siguiente secuencia (al segundo grupo de datos) $(x_{21}, \ldots, x_{2n_2}) = x_2^{(n_2)}$ con n_2 datos (donde n_2 se genera aleatoriamente). A su vez, a estos datos (junto con los de la primera etapa) se les aplica, nuevamente, la función $\psi_{\boldsymbol{n}}$ donde ahora $\boldsymbol{n} = (n_1, n_2) \in \mathcal{G}^2$, de la misma manera que en la primera etapa, para determinar: si el experimento continúa $\left(\psi_{\boldsymbol{n}}(x_1^{(n_1)}, x_2^{(n_2)}) = 0\right)$ después de esta etapa o si se detiene $\left(\psi_{\boldsymbol{n}}(x_1^{(n_1)}, x_2^{(n_2)}) = 1\right)$, y así sucesivamente.

Entonces la aplicación de la regla de paro ψ para una etapa $k \geq 1$ consiste en evaluar las respectivas funciones en el conjunto de datos respectivo:

$$\psi_{\boldsymbol{n}}\left(x_1^{(n_1)}, \ldots, x_k^{(n_k)}\right)$$

y tomar la siguiente interpretación

- se *continua* el experimento, si $\psi_{\boldsymbol{n}}\left(x_1^{(n_1)}, \ldots, x_k^{(n_k)}\right) = 0$.
- se *para* el experimento, si $\psi_{\boldsymbol{n}}\left(x_1^{(n_1)}, \ldots, x_k^{(n_k)}\right) = 1$.

Si el experimento se detiene en alguna etapa k, con los datos obtenidos de los grupos $x_1^{(n_1)}, \ldots, x_k^{(n_k)}$, significa que se han analizado k grupos y un total $n_1 + n_2 + \cdots + n_k$ datos; además la componente de la regla de paro respectiva a la etapa fue:

$$\psi_{\boldsymbol{n}}\left(x_1^{(n_1)}, \ldots, x_k^{(n_k)}\right) = 1 \quad \text{ donde } \boldsymbol{n} \in \mathcal{G}^k.$$

Ahora definiremos otra noción importante, a saber *tiempo de paro*. En el contexto ("tradicional") del análisis secuencial el *tiempo de paro* tiene su importancia porque es con base en dicho concepto que se optimiza un proceso, lo que conlleva a observaciones, tiempo y recursos menores. En este contexto dicha noción también se trabajará, aunque aquí se puede interpretar de una manera ligeramente diferente, como *el número de grupos observados* en el experimento. Por cuestiones de estilo mantendremos el concepto de tiempo de paro.

Definición 2.

La variable aleatoria denotada como τ^{ψ} se le llama **tiempo de paro** (o número de grupos observados) definida como:

$$\tau^{\psi} = \text{mín}\left\{k \geq 1 : \psi_{\boldsymbol{\nu}}\left(X_1^{(\nu_1)}, \ldots, X_k^{(\nu_k)}\right) = 1\right\}$$

El tiempo de paro es infinito si para todo $k \geq 1$,

$$\psi_{\boldsymbol{\nu}}\left(X_1^{(\nu_1)}, \ldots, X_k^{(\nu_k)}\right) = 0.$$

El tiempo paro es una noción que está asociada a una regla de paro ψ de manera biunívoca, es decir, para un tiempo de paro fijo se tiene una regla particular de paro y viceversa. De la definición anterior se observa que tan$^{\psi}$ puede ser finito o infinito. Esto último es posible cuando los grupos de datos que se vayan analizando tengan, respectivamente la componente de la regla de paro siempre igual a cero, *i. e.* $\psi_{\boldsymbol{n}} = 0$ con $\boldsymbol{n} \in \mathcal{G}^k$ para toda $k = 1, 2, \ldots$. Con base en esto y por la relación existente entre τ^{ψ} y ψ, también podemos hablar de que una regla de paro termine el experimento o no, ya que si se tiene un tiempo de paro infinito significa que la regla de paro asociada nunca termino y viceversa.

Ahora se va a encontrar la probabilidad de que $\tau^{\psi} = k$ con k finito. Como ya se mencionó, en este proceso existen dos tipos de aleatoriedad: la de los datos y la de los tamaños de los grupos. Entonces para calcular $P_{\theta}(\tau^{\psi} = k)$ debemos tomar en cuenta ambas fuentes de aleatoriedad. Calculemos primero la probabilidad condicional de una aplicación de las variables aleatorias $\nu_1, \ldots \nu_k$.

$$\begin{aligned} P_\theta^{\boldsymbol{n}}(\tau^\psi = k) &= P_\theta\Big(\tau^\psi = k \mid \nu_1 = n_1, \nu_2 = n_2, \ldots, \nu_k = n_k\Big) \\ &= P_\theta\Big(\psi_{(\boldsymbol{n}:1)}\big(X_1^{(n_1)}\big) = 0, \psi_{(\boldsymbol{n}:2)}\big(X_1^{(n_1)}, X_2^{(n_2)}\big) = 0, \ldots, \\ &\quad \psi_{(\boldsymbol{n}:k-1)}\big(X_1^{(n_1)}, X_2^{(n_2)}, \ldots X_{k-1}^{(n_{k-1})}\big) = 0, \psi_{\boldsymbol{n}}\big(X_1^{(n_1)}, X_2^{(n_2)} \ldots X_k^{(n_k)}\big) = 1\Big) \\ &= P_\theta\Big(\psi_{(\boldsymbol{n}:1)} = 0, \psi_{(\boldsymbol{n}:2)} = 0, \ldots, \psi_{(\boldsymbol{n}:k-1)} = 0, \psi_{\boldsymbol{n}} = 1\Big) \quad \text{con } \boldsymbol{n} \in \mathcal{G}^k. \end{aligned} \tag{1.1}$$

Como se ve esta probabilidad queda en términos de la regla de paro que a su vez depende de la distribución de los datos muestrales. En este punto vamos a definir dos funciones, que dependen directamente de la regla de paro y, por ende, del tiempo de paro.

Sea $\boldsymbol{n} \in \mathcal{G}^k$ definimos a

$$t_{\boldsymbol{n}}^\psi := \begin{cases} \big(1 - \psi_{(\boldsymbol{n}:1)}\big)\big(1 - \psi_{(\boldsymbol{n}:2)}\big) \cdots \big(1 - \psi_{(\boldsymbol{n}:k-1)}\big) & \text{si } k > 1; \\ 1 & \text{si } k = 1. \end{cases} \tag{1.2}$$

Cuya aplicación es la siguiente:

Si $k > 1$ y $\boldsymbol{n} = (n_1, n_2, \ldots n_k)$ entonces

$$\begin{aligned} t_{\boldsymbol{n}}^\psi &= t_{\boldsymbol{n}}^\psi\left(x_1^{(n_1)}, x_2^{(n_2)}, \ldots, x_{k-1}^{(n_{k-1})}\right) \\ &= \Big(1 - \psi_{(\boldsymbol{n}:1)}\big(x_1^{(n_1)}\big)\Big)\Big(1 - \psi_{(\boldsymbol{n}:2)}\big(x_1^{(n_1)},\ x_2^{(n_2)}\big)\Big) \cdots \\ &\qquad \cdots \Big(1 - \psi_{(\boldsymbol{n}:k-1)}\big(x_1^{(n_1)},\ x_2^{(n_2)}, \ldots,\ x_{k-1}^{(n_{k-1})}\big)\Big) \end{aligned}$$

Para el caso $k = 1$, se tendrá por definición:

$$t_n^\psi = (1 - \psi_{(\boldsymbol{n}:0)}) = 1 \tag{1.3}$$

Esta relación (1.2) está definida para cualquier vector $\boldsymbol{n}$ de dimensión $k \geq 1$.

La función $t_{\boldsymbol{n}}^{\psi}$ es una indicadora debido a que está definida mediante funciones indicadoras, $\psi_{(\boldsymbol{n}:i)}$ para $i = 1, \ldots, k-1$ con $k \geq 1$. Si $t_{\boldsymbol{n}}^{\psi} = 1$ y $\boldsymbol{n} \in \mathcal{G}^k$, entonces necesariamente nos indica que no ha habido paro hasta la etapa $k-1$.

Ahora para $\boldsymbol{n} \in \mathcal{G}^k$, definimos la función $s_{\boldsymbol{n}}^{\psi}$ [4] como:

$$\begin{aligned} s_{\boldsymbol{n}}^{\psi} &:= s_{\boldsymbol{n}}^{\psi}\left(x_1^{(n_1)}, x_2^{(n_2)}, \ldots, x_k^{(n_k)}\right) \\ &= \Big(1 - \psi_{(\boldsymbol{n}:1)}\big(x_1^{(n_1)}\big)\Big) \cdots \Big(1 - \psi_{(\boldsymbol{n}:k-1)}\big(x_1^{(n_1)},\ x_2^{(n_2)}, \ldots,\ x_{k-1}^{(n_{k-1})}\big)\Big) \cdot \\ &\quad \cdot \Big(\psi_{\boldsymbol{n}}\big(x_1^{(n_1)},\ x_2^{(n_2)}, \ldots,\ x_k^{(n_k)}\big)\Big) \\ &= \big(1 - \psi_{(\boldsymbol{n}:1)}\big)\big(1 - \psi_{(\boldsymbol{n}:2)}\big) \cdots \big(1 - \psi_{(\boldsymbol{n}:k-1)}\big)\psi_n \\ &= t_{\boldsymbol{n}}^{\psi}\, \psi_{\boldsymbol{n}}. \end{aligned} \tag{1.4}$$

La función $s_{\boldsymbol{n}}^{\psi}$ así definida es una función indicadora debido a que es el producto de las funciones indicadoras. Un punto importante del uso de la función $s_{\boldsymbol{n}}^{\psi}$ es que permite determinar el tiempo de paro de un experimento; porque para que $s_{\boldsymbol{n}}^{\psi} = 1$ es *necesario y suficiente* que el experimento termine en la etapa k, es decir

$$\boldsymbol{n} \in \mathcal{G}^k \quad \text{y} \quad s_{\boldsymbol{n}}^{\psi} = 1 \quad \Leftrightarrow \quad \tau^{\psi} = k. \tag{1.5}$$

Obsérvese que $s_{\boldsymbol{n}}^{\psi}$ es una variable aleatoria tipo *Bernoulli*, ya que sólo toma los valores cero o uno con probabilidad de éxito $p = P\big(s_{\boldsymbol{n}}^{\psi} = 1\big)$ y en consecuencia $E\big(s_{\boldsymbol{n}}^{\psi}\big) = p$.

Ahora escribiremos la probabilidad de parar en la etapa k en términos de la función $s_{\boldsymbol{n}}^{\psi}$. Para ello utilizaremos el teorema de probabilidad total y la ecuación 1.1.

[4] Donde s hace referencia a *stop*.

$$\begin{aligned} P_\theta(\tau^\psi = k) &= \sum_{\boldsymbol{n}\in\mathcal{G}^k} P_\theta^{\boldsymbol{n}}\big(\tau^\psi = k\big)\cdot P\big(\nu_1 = n_1, \nu_2 = n_2, \ldots, \nu_k = n_k\big) \\ &= \sum_{\boldsymbol{n}\in\mathcal{G}^k} P_\theta\big(\psi_{(\boldsymbol{n}:1)} = 0, \psi_{(\boldsymbol{n}:2)} = 0, \ldots, \psi_{(\boldsymbol{n}:k-1)} = 0, \psi_{\boldsymbol{n}} = 1\big)\cdot \\ &\qquad\qquad \cdot P\big(\nu_1 = n_1, \nu_2 = n_2, \ldots, \nu_k = n_k\big) \end{aligned}$$

De aquí se sigue que:

$$\begin{aligned} P_\theta(\tau^\psi = k) &= \sum_{\boldsymbol{n}\in\mathcal{G}^k} E_\theta s_{\boldsymbol{n}}^\psi\big(X_1^{(n_1)}, \ldots X_k^{(n_k)}\big)\big(P(\nu_1 = n_1)P(\nu_2 = n_2)\cdots P(\nu_k = n_k)\big) \\ &= \sum_{\boldsymbol{n}\in\mathcal{G}^k} E_\theta s_{\boldsymbol{n}}^\psi \prod_{i=1}^{k} q_{n_i}. \end{aligned}$$

Por tanto

$$P_\theta(\tau^\psi = k) = \sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}}\ E_\theta s_{\boldsymbol{n}}^\psi. \tag{1.6}$$

donde

$$q_{\boldsymbol{n}} = \prod_{i=1}^{k} q_{n_i} \tag{1.7}$$

$q_{\boldsymbol{n}}$ es el producto de las k probabilidades de los tamaños de los grupos para el vector $\boldsymbol{n} = (n_1, n_2, \ldots, n_k)$. Y $E_\theta s_{\boldsymbol{n}}^\psi$ es el valor esperado, bajo θ, de la función indicadora $s_{\boldsymbol{n}}^\psi$.

Entonces la relación (1.6) es la probabilidad, bajo el parámetro θ, de terminar el experimento cuando el tiempo de paro es igual a k. [5] Esta probabilidad esta suponiendo que la regla de paro ψ asociada al tiempo de paro τ^ψ, termina el experimento en la etapa k. Ahora se analizará qué reglas de paro tienen una probabilidad alta de terminar antes de la etapa k.

[5] O la probabilidad observar k grupos de tamaños $\nu_1, \nu_2, \ldots, \nu_k$ respectivamente.

1.3.1. Reglas de paro que terminan con probabilidad uno

La ecuación (1.6) nos da la probabilidad de que se termine un experimento con un tiempo de paro igual a k, pero es importante saber cuándo esta probabilidad está bien definida. En particular se requieren de procesos en los que el tiempo de paro tengan un valor no infinito, esto es que se satisfaga la ecuación $P_\theta\big(\tau^\psi = \infty\big) = 0$.

Definición 3.

Sea Ψ_θ la familia de funciones, definida como:

$$\Psi_\theta = \Big\{\psi \text{ regla de paro : } \; P_\theta\big(\tau^\psi = \infty\big) = 0\Big\}$$

La clase Ψ_θ contiene a las funciones de paro ψ para la cuales no sucede que su respectivo tiempo de paro sea infinito, bajo cierto parámetro θ, o dicho de otra manera *las reglas de paro que pertenecen a esta clase terminan el experimento con probabilidad uno*, bajo el parámetro θ. Esto debido al siguiente argumento.

Los eventos $\{\tau^\psi = 1\}$, $\{\tau^\psi = 2\}, \ldots, \{\tau^\psi = k\}, \ldots, \{\tau^\psi = \infty\}$, son mutuamente excluyentes, entonces la unión de ellos conforman todo el espacio Ω de probabilidad, esto es:

$$\bigcup_{k=1}^{\infty}\{\tau^\psi = k\} \cup \{\tau^\psi = \infty\} = \Omega$$

Tomando la probabilidad bajo el parámetro θ, se tiene que:

$$\sum_{k=1}^{\infty} P_\theta(\tau^\psi = k) + P_\theta\big(\tau^\psi = \infty\big) = 1 \tag{1.8}$$

De aquí se sigue que una regla de paro $\psi \in \Psi_\theta$ si y sólo si

$$\sum_{k=1}^{\infty} P_\theta(\tau^\psi = k) = 1. \tag{1.9}$$

O de manera equivalente:

$$P_\theta(\tau^\psi < \infty) = 1. \tag{1.10}$$

Lo cual nos indica que si una regla de paro pertenece a la clase Ψ_θ (definición 3) entonces dicha regla ψ *tiene un tiempo de paro finito con probabilidad uno* [6], bajo θ. Otra propiedad es que el número de grupos observados en el experimento no crece indefinidamente debido a que:

Dado N un número natural cualesquiera, definamos el conjunto:

$$A_N = \left\{\tau^\psi \geq N\right\}.$$

y dado que

$$\tau^\psi \geq N+1 \;\Rightarrow\; \tau^\psi \geq N$$

entonces:

$$A_{N+1} \;\subset\; A_N$$

Por tanto se obtiene la sucesión de conjuntos A_N con la propiedad:

$$\ldots A_{N+2} \subset A_{N+1} \subset A_N.$$

Por definición de límite se tiene:

$$\lim_{N\to\infty} A_N \;=\; \bigcap_{N\geq 1} A_N = \{\tau^\psi = \infty\} \tag{1.11}$$

[6] El número de grupos observados en el experimento será finito.

Ahora por continuidad y dado que $\psi \in \Psi_\theta$:

$$\lim_{N\to\infty} P_\theta(A_N) = P_\theta\left(\lim_{N\to\infty} A_N\right)$$

$$\Leftrightarrow \qquad \lim_{N\to\infty} P_\theta\left(\tau^\psi \geq N\right) = P_\theta\left(\tau^\psi = \infty\right) = 0 \tag{1.12}$$

Esta última relación nos indica que la probabilidad de que el tiempo de paro crezca indefinidamente será nula *siempre y cuando* la regla de paro $\psi \in \Psi_\theta$.

En general, en este trabajo nos interesan la reglas de paro que conlleven tiempos de paro finitos, debido a que ello es más aplicable. Entonces

Dado el parámetro θ, para la clase Ψ_θ se tiene que: $\psi \in \Psi_\theta$ si y sólo si:

1.- ψ termina el experimento con probabilidad uno $\left(P_\theta(\tau^\psi < \infty) = 1\right)$;

2.- Y la probabilidad de que el tiempo de paro τ^ψ crezca indefinidamente es nula, $\left(\lim_{N\to\infty} P_\theta\left(\tau^\psi \geq N\right) = 0\right)$.

Puede observarse que las dos propiedades anteriores son equivalentes y ellas conllevan pertenecer a la clase Ψ_θ.

Ahora, vamos a tomar en cuenta las propiedades anteriores para el cálculo de la probabilidad de que el número de grupos observados sea k. Supongamos que las reglas de paro ψ en el experimento pertenecen a la clase Ψ_θ, entonces al sustituir $P_\theta(\tau^\psi = k)$ en la ecuación (1.9) resulta que:

$$\sum_{k=1}^{\infty} P_\theta(\tau^\psi = k) = 1 \quad \Leftrightarrow \quad \sum_{k=1}^{\infty} \sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}} E_\theta s_{\boldsymbol{n}}^\psi = 1. \tag{1.13}$$

Entonces el experimento puede terminar en cualquier etapa finita k ($k = 1, 2, 3, \dots$).

En particular, a lo largo del trabajo, vamos a trabajar con la clase Ψ_{θ_0} donde las reglas de paro terminan el experimento con probabilidad uno bajo θ_0, relacionada a nuestra hipótesis H_0. [7] Al final del trabajo también se trabajará con la clase Ψ_{θ_1}.

1.4. Reglas de decisión terminal

Continuando con nuestro experimento es conveniente ahora preguntarse por la desición final de estar a favor de H_0 o de H_1, con base en el tiempo de paro τ^{ψ} obtenido. Para ello se necesita una regla que nos permita decidir sobre la hipótesis en cuestión. Dicha regla será llamada *la regla de decisión terminal* denotada con ϕ. Esta regla, al igual que la de paro, será una función indicadora.

Definición 4.

Una *regla de decisión* ϕ es una familia de funciones indicadoras

$$\left\{\phi_{\boldsymbol{n}} : \mathcal{X}^{n_1} \times \mathcal{X}^{n_2} \times \cdots \times \mathcal{X}^{n_k} \to \{0,1\}, \text{ con } \boldsymbol{n} \in \mathcal{G}^k \text{ y } k \geq 1\right\}$$

Donde dada una realización del vector $\boldsymbol{\nu} \in \mathcal{G}^k$ entonces si $\boldsymbol{n} \in \mathcal{G}^k$, la aplicación de la regla de decisión ϕ para una etapa k consiste en evaluar la función:

[7] Esto debido a que primero se trabajará con el parámetro θ_0 y después las mismas propiedades pero con θ_1.

$$\phi_{\boldsymbol{n}} = \phi_{\boldsymbol{n}}\left(x_1^{(n_1)}, x_2^{(n_2)}, \ldots, x_k^{(n_k)}\right).$$

en los grupos de los datos muestrales y tomar la siguiente decisión:

- se *acepta* la hipótesis H_0, si $\phi_{\boldsymbol{n}}\left(x_1^{(n_1)}, \ldots, x_k^{(n_k)}\right) = 0$.
- se *rechaza* la hipótesis H_0, a favor de H_1, si $\phi_{\boldsymbol{n}}\left(x_1^{(n_1)}, \ldots, x_k^{(n_k)}\right) = 1$.

Podemos observar que para aplicar alguna de las componentes de la regla de decisión es necesario que el experimento haya terminado. Con base en el tiempo de paro observado, uno puede preguntarse por la decisión final para la prueba de hipótesis correspondiente. Entonces este contexto nos lleva a la siguiente definición.

1.5. Prueba estadística secuencial

Ahora tomando en cuenta que en un experimento secuencial se tiene una regla de paro ψ y una regla de decisión ϕ se tiene la siguiente definición:

Definición 5.

Una *prueba estadística secuencial* es una pareja (ψ, ϕ) compuesta por una regla de paro ψ y una regla de decisión terminal ϕ.

Esta pareja nos proporciona una manera de proceder en el análisis de los datos, para llegar a una conclusión con respecto a nuestras hipótesis. Aunque, recordemos, que se está trabajando con dos aleatoriedades (la de los datos y la de los tamaños de los grupos), por tanto el proceso de la prueba estadística podría conllevar a tomar la decisión incorrecta.

1.5.1. Probabilidades de error

Una de las maneras en que una prueba de hipótesis "mide" un equívoco es utilizando las probabilidades de error. Los tipos de errores que se toman en cuenta en estas pruebas son conocidos como: **error tipo I** y **error tipo II**. El primero sucede cuando siendo la hipótesis H_0 verdadera, la decisión final es rechazarla. El error tipo II, al contrario, se da cuando siendo falsa la hipótesis H_0, la decisión final es aceptarla.

Para tener una calidad en la inferencia estadística se debe tener en cuenta que dichos errores no sucedan con frecuencia. Para esto es necesario "controlarlos" mediante sus probabilidades. Entonces denotemos la *probabilidad* de error tipo I como $\alpha(\psi, \phi)$ y la probabilidad del error tipo II como $\beta(\psi, \phi)$ donde:

$$\alpha(\psi, \phi) = P_{\theta_0}(\text{rechazar } H_0) \;\text{ y }\; \beta(\psi, \phi) = P_{\theta_1}(\text{aceptar } H_0).$$

Ambas probabilidades de error están en dependencia de las funciones de paro y decisión terminal (ψ, ϕ respectivamente), debido a que éstas son necesarias durante el proceso de la prueba de hipótesis. También nótese que cada probabilidad de error está asociada, respectivamente, a la hipótesis que se rechaza. Ahora calculemos estas probabilidades.

$$\begin{aligned}
\alpha(\psi,\phi) &= P_{\theta_0}(\text{rechazar } H_0) = \sum_{k=1}^{\infty} P_{\theta_0}(\text{rechazar } H_0 \text{ en la etapa } k) \\
&= \sum_{k=1}^{\infty} \sum_{\boldsymbol{n}\in\mathcal{G}^k} P_{\theta_0}\big(\text{rechazar } H_0 \text{ en la etapa } k \text{ dados } \nu_1 = n_1,\ldots,\nu_k = n_k\big) \\
&\qquad\qquad \cdot P(\nu_1 = n_1,\ldots,\nu_k = n_k) \\
&= \sum_{k=1}^{\infty} \sum_{\boldsymbol{n}\in\mathcal{G}^k} P_{\theta_0}\big(\phi_{\boldsymbol{n}} = 1,\ \tau^{\psi} = k\big)\cdot q_{\boldsymbol{n}} = \sum_{k=1}^{\infty} \sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}} P_{\theta_0}\big(\phi_{\boldsymbol{n}} = 1,\ \tau^{\psi} = k\big).
\end{aligned}$$

Sin embargo, tengamos presente el hecho dado en (1.5). Por tanto

$$\begin{aligned}
\alpha(\psi,\phi) &= \sum_{k=1}^{\infty} \sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}}\, P_{\theta_0}\big(\phi_{\boldsymbol{n}} = 1,\ s_{\boldsymbol{n}}^{\psi} = 1\big) \\
&= \sum_{k=1}^{\infty} \sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}}\, P_{\theta_0}\big(\phi_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi} = 1\big) \\
&= \sum_{k=1}^{\infty} \sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}}\, \big(E_{\theta_0} \phi_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi}\big). \qquad (1.14)
\end{aligned}$$

Ahora calculemos la probabilidad del error tipo II, $\beta(\psi,\phi)$.

$$\begin{aligned}
\beta(\psi,\phi) &= P_{\theta_1}(\text{aceptar } H_0) = \sum_{k=1}^{\infty} P_{\theta_1}(\text{aceptar } H_0 \text{ en la etapa } k) \\
&= \sum_{k=1}^{\infty} \sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}} P_{\theta_1}\big(\phi_{\boldsymbol{n}} = 0,\ \tau^{\psi} = k\big) \\
&= \sum_{k=1}^{\infty} \sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}}\, P_{\theta_1}\big(\phi_{\boldsymbol{n}} = 0,\ s_{\boldsymbol{n}}^{\psi} = 1\big) \\
&= \sum_{k=1}^{\infty} \sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}}\, P_{\theta_1}\big((1-\phi_{\boldsymbol{n}}) s_{\boldsymbol{n}}^{\psi} = 1\big) \\
&= \sum_{k=1}^{\infty} \sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}}\, \Big(E_{\theta_1}\big(1-\phi_{\boldsymbol{n}}\big) s_{\boldsymbol{n}}^{\psi}\Big) \qquad (1.15)
\end{aligned}$$

Se sabe, por la definición 2, que el tiempo de paro (τ^{ψ}) en un experimento es una variable aleatoria y dicho valor puede ser estimado con su valor promedio bajo el parámetro θ_0 o bajo θ_1. Entonces denotemos dichos promedios por medio de

$$\mathcal{N}(\theta_0;\psi) = E_{\theta_0}\tau^{\psi} \quad \text{y} \quad \mathcal{N}(\theta_1;\psi) = E_{\theta_1}\tau^{\psi}$$

Estos promedios son utilizados en trabajos conocidos del análisis secuencial. Los correspondientes promedios del tiempo de paro ($\mathcal{N}(\theta;\psi)$) no dependen de la función de decisión ϕ, sólo de la regla paro ψ, debido a que la variable τ^{ψ} sólo depende de dicha función (véase definición 2) por ello es que ésta sólo tiene marcado la dependencia de ψ.

Ahora al calcular el valor esperado del tiempo de paro para un parámetro θ cualquiera, con $\psi \in \Psi_{\theta}$ se tiene:

$$\begin{aligned}
\mathcal{N}(\theta;\psi) &= E_{\theta}\tau^{\psi} = \sum_{k=1}^{\infty} k \ P_{\theta}(\tau^{\psi} = k) \\
&= \sum_{k=1}^{\infty} \sum_{\boldsymbol{n}\in\mathcal{G}^k} k \ q_{\boldsymbol{n}} \ E_{\theta} s_{\boldsymbol{n}}^{\psi}. \qquad (1.16)
\end{aligned}$$

Hasta este momento se tiene que: dado un experimento en el cual las observaciones ($x_k^{(n_k)}$) vienen en grupos de tamaño (n_k) (ambos aleatorios), se analizan dichas observaciones mediante una función de paro (ψ) y una función de decisión (ϕ) para determinar, con respecto a ciertos parámetros ($\theta_0, \ \theta_1$), cuál de ellos es el indicado para la distribución de probabilidad de los datos, bajo el hecho de que $\psi \in \Psi_{\theta}$. Esto se realiza mediante una prueba de hipótesis secuencial, en la cual se toman en cuenta las respectivas probabilidades de error ($\alpha, \ \beta$).

1.5.2. Costo promedio

Dado un experimento estadístico y una prueba secuencial (ψ, ϕ) se busca que dicha prueba conlleve a un punto óptimo, en este contexto, este punto será el costo del experimento[8], ya que nos interesa tener un costo de implementación mínima. Para encontrar la prueba optima introducimos los respectivos costos de cada grupo de observaciones.

Sea n_k el tamaño de un grupo de observaciones, el costo que se tiene que pagar por los n_k datos será c_{n_k}, para cualquier $k = 1, 2, \ldots$. Entonces sea el costo $c_{n_k} : \mathcal{G} \to \mathbb{R}^+$ una función no negativa que depende del número de observaciones tratadas en cada grupo, más recuérdese que dicho número es aleatorio (correspondiente a ν_k, el tamaño del grupo), entonces podemos decir que se tiene una sucesión de variables aleatorias correspondientes a los costos de cada grupo, esto es, se tendrá la sucesión de los respectivos costos por cada grupo: $c_{\nu_1}, c_{\nu_2}, \ldots c_{\nu_k}, \ldots$. Donde para una realización dada $n_1, n_2, \ldots, n_k$ de tamaños de k grupos, se tiene $c_{n_1}, c_{n_2}, \ldots, c_{n_k}$ sus respectivos costos, y c_{n_k} es cualquier función no negativa, pero fija en el experimento. Entonces

El *costo promedio de un grupo* se denota:

$$\bar{c} = \sum_{m \in \mathcal{G}} q_m c_m \tag{1.17}$$

donde $P(\nu = m) = q_m$.

Como las variables aleatorias ν_k son finitas, independientes e idénticamente distribuidas, entonces las variables c_{ν_k}, también tienen las mismas propiedades. Por tanto el valor esperado del costo de un grupo es el mismo para cualquier grupo. Además podemos aseverar

[8]En el contexto clásico el punto de optimalidad es el promedio del tiempo de paro $\mathcal{N}(\theta; \psi)$.

que dicho costo promedio $\overline{c}$ es finito por ser la suma finita de valores finitos. Por tanto $\overline{c} < \infty$. Además como los c_ν son funciones no negativas entonces se tiene que $0 \leq \overline{c}$. Sin embargo si $\overline{c} = 0$, significa que $c_\nu = 0$ para toda ν, más este caso no lo vamos a tener en cuenta debido a que no es de nuestro interés tener un experimento en el cual el costo para todo grupo de observaciones es nulo. Es decir, si no se tiene un costo para cada grupo [9], entonces el procedimiento óptimo es nunca terminar. De lo cual, se tendrá que $0 < \overline{c} < \infty$. También es importante destacar que el costo c_{n_k} de cada grupo de observaciones, no necesariamente es una función proporcional al número de observaciones n_k de cada grupo. Dicha relación, en general puede ser cualquier función no negativa.

Ahora se tiene *el costo total de k etapas*: dado $\boldsymbol{n} \in \mathcal{G}^k$,

$$c(\boldsymbol{n}) = c_{n_1} + \cdots + c_{n_k}. \tag{1.18}$$

Y el costo total cuando el experimento no termina:

$$c\big(\boldsymbol{\nu}^\infty\big) = \lim_{k\to\infty} \big(c_{\nu_1} + \cdots + c_{\nu_k}\big). \tag{1.19}$$

Con base en estas notaciones se determina el *costo total esperado* del experimento:

$$\begin{aligned} K(\theta;\psi) &= E_\theta\{\text{Costo total}\} \\ &= \sum_{k=1}^{\infty} E_\theta\Big(\big(c_{\nu_1} + c_{\nu_2} + \cdots + c_{\nu_k}\big)\ I_{\{\tau^\psi = k\}}\Big) + \ E_\theta\Big(\big(c_{\nu_1} + c_{\nu_2} + \cdots\big)\ I_{\{\tau^\psi = \infty\}}\Big) \end{aligned} \tag{1.20}$$

Obsérvese que el último termino de la expresión toma en cuenta cuando el experimento no termina.

[9] O no cuesta nada.

Cabe mencionar que otra manera de trabajar el costo promedio del experimento sería el utilizar la identidad de Wald, de la siguiente forma:

$$\begin{aligned} K(\theta;\psi) &= E_\theta\Big(c_{\nu_1} + c_{\nu_2} + \cdots + c_{\nu_{\tau^\psi}}\Big) \\ &= E_\theta\Big(\sum_{k=1}^{\tau^\psi} c_{\nu_k}\Big) \\ &= E_\theta c_{\nu_1} E_\theta \tau^\psi = \overline{c}\ E_\theta \tau^\psi \end{aligned} \tag{1.21}$$

Sin embargo, para la aplicación de la identidad de Wald en lo anterior, se necesita que el evento $\{\tau^\psi \geq n\}$ dependa de las variables aleatorias ν_i con $1 \leq \nu_i < n$, mas nuestro tiempo de paro también depende de las observaciones, por lo cual no se tiene la certeza de la aplicación de la identidad. Este camino es un punto interesante de investigación futura, ya que si se encuentra una prueba secuencial que conlleve a un costo promedio finito $\big(K(\theta;\psi) < \infty\big)$, entonces se podrá asegurar que el promedio del tiempo de paro del experimento también será finito ($E_\theta \tau^\psi < \infty$).

Regresando al análisis de la ecuación 1.20, se tiene:

$$\begin{aligned} K(\theta;\psi) &= \sum_{k=1}^{\infty} \sum_{\boldsymbol{n}\in\mathcal{G}^k} E_\theta\Big(\big(c_{\nu_1} + \cdots + c_{\nu_k}\big) I_{\{\tau^\psi = k\}} \mid \nu_1 = n_1, \ldots, \nu_k = n_k\Big) \\ &\qquad P\big(\nu_1 = n_1, \ldots, \nu_k = n_k\big)\ +\ E_\theta\Big(c(\nu^\infty)\ I_{\{\tau^\psi = \infty\}}\Big) \\ &= \sum_{k=1}^{\infty} \sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}} E_\theta\Big(\big(c_{n_1} + c_{n_2} + \cdots + c_{n_k}\big)\ s_{\boldsymbol{n}}^\psi\Big) +\ E_\theta\Big(c(\nu^\infty)\ I_{\{\tau^\psi = \infty\}}\Big) \\ &= \sum_{k=1}^{\infty} \sum_{\boldsymbol{n}\in\mathcal{G}^k} \big(c_{n_1} + \cdots + c_{n_k}\big) q_{\boldsymbol{n}}\ E_\theta s_{\boldsymbol{n}}^\psi\ +\ E_\theta\Big(c(\nu^\infty)\ I_{\{\tau^\psi = \infty\}}\Big). \end{aligned}$$

Entonces

$$K(\theta;\psi) = \sum_{k=1}^{\infty} \sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}}\ c(\boldsymbol{n})\ E_\theta s_{\boldsymbol{n}}^\psi\ +\ E_\theta\Big(c(\nu^\infty)\ I_{\{\tau^\psi = \infty\}}\Big). \tag{1.22}$$

En el caso cuando se trabaja con una regla de paro que termina el experimento $\big(\psi \in \Psi_\theta\big)$, el segundo término del costo no aparece por las propiedades de la regla de paro. Entonces:

$$\psi \in \Psi_\theta \quad \Rightarrow \quad K(\theta;\psi) \;=\; \sum_{k=1}^{\infty} \sum_{\boldsymbol{n} \in \mathcal{G}^k} q_{\boldsymbol{n}}\ c(\boldsymbol{n})\ E_\theta s_{\boldsymbol{n}}^{\psi}. \tag{1.23}$$

En este contexto, analicemos algunos casos de los costos de un grupo:

a) Cuando el **costo es fijo** para todos los grupos, independientemente de su tamaño, por ejemplo se puede tomar un costo unitario $c_{n_k} = 1$ o un costo fijo $c_{n_k} = c$ para cualquier tamaño n_k . Para el primer caso, la suma de los costos de los k grupos sería el mismo valor k, *i. e.* $c_{n_1} + \ldots + c_{n_k} = k$. Ahora, si se sustituye este valor en el costo total promedio $K(\theta;\psi)$ (1.22) entonces dicho costo coincide exactamente con el número promedio de grupos observados $\mathcal{N}(\theta;\psi)$ (véase (1.16)). Por lo cual se tendría, en este caso, que $K(\theta;\psi) = \mathcal{N}(\theta;\psi)$.

Para el caso en el que se tiene el mismo costo $c_{n_k} = c$ con $c > 0$, la suma correspondiente sería $c_{n_1} + \cdots + c_{n_k} = k \cdot c$ y por tanto $K(\theta;\psi) = c\mathcal{N}(\theta;\psi)$. Cabe mencionar que este caso en particular es el que se trabaja cuándo los grupos son de tamaño uno, esto es se van analizando las observaciones de una por una ([12], [11]).

b) Cuando el costo es directamente **proporcional** a la cantidad de datos en cada grupo, es decir si $c_{n_k} = c\, n_k$ para cualquier n_k. En este caso la suma correspondiente sería $c_{n_1} + \cdots + c_{n_k} = c(n_1 + n_2 + \cdots + n_k)$. Con lo cual el costo promedio total del experimento $K(\theta;\psi)$ será proporcional al número total de datos observados.

En general, no se tiene una relación específica para el costo de cada grupo de observaciones, éste podría tener cualquier estructura, como por ejemplo la forma $c_{n_k} = \kappa\, n_k + l$, cuya interpretación podría ser: κ un costo unitario por unidad más un costo extra por lote (l). Puede notarse que nuestra forma del costo promedio total engloba los casos particulares anteriormente descritos.

Por otro lado, en el experimento se podría tener una regla de paro ψ que no lo termine. Para estos casos analicemos el costo total del experimento.
Por la ley de grandes números para la sucesión creciente $c_{\nu_1}, c_{\nu_2}, \ldots$, se tiene que:

$$P\left(\lim_{k\to\infty}\left(\frac{c_{\nu_1}+\cdots+c_{\nu_k}}{k}\right)=\overline{c}\right)=1$$

$$\Rightarrow \quad P\left(\lim_{k\to\infty}\left(c_{\nu_1}+\cdots+c_{\nu_k}\right)=\infty\right)=1 \quad \text{si} \quad \overline{c}>0 \tag{1.24}$$

De lo cual, si ψ es una regla de paro que no termina el experimento,

$$\begin{aligned} K(\theta;\psi) &= E_\theta\Big(c(\nu^\infty)\ I_{\{\tau^\psi=\infty\}}\Big) = E_\theta\Big(\lim_{k\to\infty}\left(c_{\nu_1}+\cdots+c_{\nu_k}\right)\ I_{\{\tau^\psi=\infty\}}\Big) \\ &= \lim_{k\to\infty}\left(c_{\nu_1}+\cdots+c_{\nu_k}\right)P_\theta\left(\tau^\psi=\infty\right) \end{aligned} \tag{1.25}$$

En general, si ψ es una regla de paro que no termina el experimento, entonces el costo promedio del experimento para dicha regla no es finito, a menos que $\overline{c}=0$.

$$\psi \notin \Psi_\theta \quad \Rightarrow \quad K(\psi;\theta)=\infty \tag{1.26}$$

En particular, vamos a estar trabajando con el parámetro θ_0, entonces se tendrá $K(\theta_0;\psi)$; y a menos que se especifique lo contrario, tomaremos $K(\psi)=K(\theta_0;\psi)$, esto es:

$$K(\psi) \;=\; \sum_{k=1}^{\infty}\sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}}\ c(\boldsymbol{n})\ E_{\theta_0}s_{\boldsymbol{n}}^{\psi} \tag{1.27}$$

para $\psi\in\Psi_{\theta_0}$.

CAPÍTULO 2

OPTIMALIDAD DE LAS PRUEBAS SECUENCIALES CON GRUPOS DE TAMAÑO ALEATORIO

2.1. Planteamiento del problema: optimalidad

En el contexto del análisis secuencial clásico, en donde los datos se van analizando de uno en uno, se sabe que el costo promedio total del experimento secuencial es *proporcional* al número de datos observados (véase [6],[13]). Uno de los problemas en dicho contexto es minimizar el costo total del experimento, y se llega a la forma de la prueba que minimiza el tiempo de paro y también minimiza el número promedio de observaciones en el experimento (véase [13]).

Sin embargo en nuestro contexto, y de manera más general, se busca optimizar el costo promedio del experimento. Como se mencionó anteriormente el primer problema es un caso particular de nuestra investigación. Entonces formalicemos nuestro problema de investigación.

Sean $\Delta(\alpha, \beta)$ la clase de todas las pruebas secuenciales (ψ, ϕ) tales que se satisfacen las siguientes desigualdades:

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

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

El problema de investigación consiste en encontrar una prueba $(\psi, \phi) \in \Delta(\alpha, \beta)$ tal que su costo correspondiente $K(\theta_0; \psi)$ es mínimo en $\Delta(\alpha, \beta)$. Esto significa que en un experimento secuencial existe una prueba secuencial con ciertas cotas dadas (relacionadas con los errores tipo I y tipo II) tal que el costo promedio de dicho experimento es mínimo. Entonces tenemos que minimizar la función (1.27) con ciertas restricciones, esto es

$$\text{mín } K(\psi) \quad \text{restringido a} \quad \alpha(\psi, \phi) \leq \alpha \quad \text{y} \quad \beta(\psi, \phi) \leq \beta. \tag{2.1}$$

$$\text{donde } \ \alpha, \beta \in [0, 1) \text{ constantes fijas.}$$

Aquí tenemos un problema de optimización con restricciones y este lo vamos a resolver utilizando el método variacional de Lagrange. Como se sabe este método nos permite expresar a nuestra función a optimizar y a sus restricciones como una sola función, entonces sea la función de Lagrange definida como:

$$L(\psi, \phi) = K(\psi) + \lambda_0\ \alpha(\psi, \phi) + \lambda_1\ \beta(\psi, \phi). \tag{2.2}$$

donde $\lambda_0 \geq 0$ y $\lambda_1 \geq 0$ son dos constantes cualesquiera conocidas como los multiplicadores de Lagrange.

Para resolver el problema de minimización de la función de Lagrange, sobre las pruebas secuenciales, vamos a buscar la estructura de una prueba secuencial (ψ, ϕ) tal que alcance la cota mínima (si ésta existe) de la función (2.2)

2.2. Resolución del problema de optimalidad

El siguiente teorema representa la reducción del problema con restricciones a un problema sin restricciones en la que propiamente consiste el método de multiplicadores de Lagrange.

Teorema 1.

Sea (ψ^*, ϕ^*) una prueba secuencial tal que:

$$\begin{aligned} &\textbf{\textit{a)}}\quad \alpha(\psi^*, \phi^*) = \alpha \quad y \quad \beta(\psi^*, \phi^*) = \beta; \text{ y} \\ &\textbf{\textit{b)}}\quad \forall\ (\psi, \phi) \quad L(\psi, \phi) \geq L(\psi^*, \phi^*), \end{aligned} \tag{2.3}$$

Entonces para cualquier prueba (ψ, ϕ) que satisface

$$\textbf{\textit{c)}}\qquad \alpha(\psi, \phi) \leq \alpha \quad y \quad \beta(\psi, \phi) \leq \beta \tag{2.4}$$

$$\textit{se tiene que} \quad K(\psi) \geq K(\psi^*). \tag{2.5}$$

La desigualdad en (2.5) es estricta si alguna de las desigualdades en (2.4) lo es.

Nótese que es obvio que (2.3) (c) garantiza que $(\psi^*, \phi^*) \in \Delta(\alpha, \beta)$, así que (2.5) representa la optimalidad de (ψ^*, ϕ^*) en la clase $\Delta(\alpha, \beta)$ desde el punto de vista del costo promedio del experimento.

Demostración.

Sea (ψ^*, ϕ^*) una prueba que satisface (2.3) (a)-(b); además sea (ψ, ϕ) cualquier otra prueba que satisface que $\alpha(\psi, \phi) \leq \alpha$ y $\beta(\psi, \phi) \leq \beta$. Para $\lambda_0 \geq 0$ y $\lambda_1 \geq 0$ se satisface la desigualdad:

$$\lambda_0\alpha + \lambda_1\beta \geq \lambda_0\alpha(\psi, \phi) + \lambda_1\beta(\psi, \phi).$$

$$\begin{aligned}
\Rightarrow \quad K(\psi) + \lambda_0\alpha + \lambda_1\beta &\geq K(\psi) + \lambda_0\alpha(\psi, \phi) + \lambda_1\beta(\psi, \phi) \\
&= L(\psi, \phi) \\
&\geq L(\psi^*, \phi^*) \\
&= K(\psi^*) + \lambda_0\alpha(\psi^*, \delta^*) + \lambda_1\beta(\psi^*, \delta^*) \\
&= K(\psi^*) + \lambda_0\alpha + \lambda_1\beta. \qquad (2.6)
\end{aligned}$$

Por tanto de la cadena anterior se tiene que

$$\begin{aligned}
K(\psi) + \lambda_0\alpha + \lambda_1\beta &\geq K(\psi^*) + \lambda_0\alpha + \lambda_1\beta \\
\Rightarrow \quad K(\psi) &\geq K(\psi^*). \qquad (2.7)
\end{aligned}$$

lo que demuestra el teorema. Cabe señalar que si ahora $K(\psi) = K(\psi^*)$ para alguna prueba (ψ, ϕ), entonces *todas* las desigualdades anteriores a (2.6) en realidad serán igualdades, por lo que $\alpha(\psi, \phi) = \alpha$ *y* $\beta(\psi, \phi) = \beta$, y esto concluye la demostración. ∎

El teorema 1 nos permite dedicarnos a la solución del problema de minimización de la función de Langrange $L(\psi, \phi)$ $\big($ver la condición (2.3)(b)$\big)$. Como hay dos constantes arbitrarias (λ_0, λ_1) en la función de Lagrange, se puede esperar que variando éstas uno pueda cumplir con las condiciones (2.3)(a). De cierta forma, estas condiciones son automáticas, ya que una vez que exista una prueba (ψ^*, ϕ^*) que satisface (2.3)-(b), la condición (2.3)-(a) se cumple si tomamos $\alpha = \alpha(\psi^*, \phi^*)$ y $\beta = \beta(\psi^*, \phi^*)$. De hecho, es precisamente el sentido que se le da a la optimalidad de la prueba SPRT (véase [20]).

2.2.1. Reducción al problema del paro óptimo

Del Teorema 1 vemos que para hallar la forma de la prueba óptima necesitamos minimizar la función de Lagrange $L(\psi, \phi)$. En esta sección este problema recibe una resolución parcial: resulta posible dar una regla de decisión ϕ^* óptima en el sentido de que para cualquier otra regla de decisión ϕ,

$$L(\psi, \phi) \geq L(\psi, \phi^*).$$

Esto es, el problema de optimización de la función de Lagrange (2.2) consistirá en encontrar primero la optimización de $L(\psi, \phi)$ para cualquier regla de paro ψ

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

Después de ello, se encontrará la función de Lagrange óptima ahora para la función $L(\psi, \phi^*)$. Con estas reglas de paro ψ^* y de decisión ϕ^* fijas, se tendrá la mínima cota de la función de Lagrange $L(\psi^*, \phi^*)$ para la prueba secuencial (ψ^*, ϕ^*).

Para introducir ϕ^* tomemos la siguiente notación.

Sea $\boldsymbol{n} \in \mathcal{G}^k$ y se define la función:

$$\begin{aligned}
f_i^{\boldsymbol{n}} &= f_i^{\boldsymbol{n}}\left(x_1^{(n_1)}, x_2^{(n_2)}, \ldots, x_k^{(n_k)}\right) \quad i = 0, 1. \\
&= f_i^{\boldsymbol{n}}\left(\underbrace{x_{11}, x_{12}, \ldots, x_{1n_1}}_{datos\ de\ x_1^{(n_1)}}, \underbrace{x_{21}, x_{22}, \ldots, x_{2n_2}}_{datos\ de\ x_2^{(n_2)}}, \ldots, \underbrace{x_{k1}, x_{k2}, \ldots, x_{kn_k}}_{datos\ de\ x_k^{(n_k)}}\right) \\
&= \prod_{m=1}^{k} \prod_{j=1}^{n_m} f_{\theta_i}(x_{mj}) \quad \text{con } i = 0, 1.
\end{aligned} \tag{2.8}$$

La función $f_i^{\boldsymbol{n}}$ con $i = 0, 1$ es el producto de probabilidad conjunta de los datos de cada grupo. Más es importante señalar que pueden existir grupos que **no** tengan ninguna observación, esto es, el tamaño de estos sea cero. Por lo cual su correspondiente vector será de la forma $x_i^{(0)} = ()$. Para casos como este se toma por definición a su producto de probabilidad conjunta como 1, esto es $\prod_{j=1}^{0} \equiv 1$. Con esto nuestra ecuación (2.8) estará bien definida.

Entonces, la función $f_i^{\boldsymbol{n}}$ tiene dos posibilidades, a saber $f_0^{\boldsymbol{n}}$ y f_1^{n} respectivamente para los parámetros θ_0 y θ_1, con $\boldsymbol{n} \in \mathcal{G}^k$.

Ahora sea

$$\phi_{\boldsymbol{n}}^* = I_{\left\{\lambda_0 f_0^{\boldsymbol{n}} \leq \lambda_1 f_1^{\boldsymbol{n}}\right\}} = \begin{cases} 1, & \text{si } \lambda_0 f_0^{\boldsymbol{n}} \leq \lambda_1 f_1^{\boldsymbol{n}}, \\ 0, & \text{en otro caso.} \end{cases} \tag{2.9}$$

La relación anterior $\phi_{\boldsymbol{n}}^*$ corresponde a las componentes para aplicar a cada vector $\boldsymbol{n} = (n_1, n_2, \ldots, n_k)$. Entonces, la regla de decisión general está formada por la familia,

$$\phi^* = \left\{\phi_{\boldsymbol{n}}^*, \quad \boldsymbol{n} \in \mathcal{G}^k, \quad k \geq 1\right\}.$$

La ecuación (2.9) muestra la estructura de los elementos de la regla de decisión que vamos a utilizar. A continuación nos dedicaremos a mostrar que dicha regla es en realidad óptima. Para lo cual presentaremos un sencillo lema que nos ayudará a la demostración de varios teoremas importantes.

Lema 1.

Sea x una variable con un número finito de valores; y sean $F_1(x)$ y $F_2(x)$ cualesquiera funciones de x. Entonces para cualesquiera funciones $\phi(x), \chi(x)$ con $0 \leq \phi(x), \chi(x) \leq 1$, se tiene

$$\sum_x \chi(x)\Big(\phi(x)F_1(x)+\big(1-\phi(x)\big)F_2(x)\Big) \geq \sum_x \chi(x) \min\big\{F_1(x), F_2(x)\big\}. \tag{2.10}$$

En el caso en el que $\phi(x) = I_{\{F_1 \leq F_2\}}(x)$ entonces (2.10) se convierte en igualdad.

Demostración.

Sean $F_1(x)$, $F_2(x)$ y $\phi(x), \chi(x)$ cualesquiera funciones de x con $0 \leq \phi(x), \chi(x) \leq 1$. Sabemos que $F_1(x) \geq \min\{F_1(x), F_2(x)\}$ y $F_2(x) \geq \min\{F_1(x), F_2(x)\}$.

Entonces de ambas desigualdades se tiene:

$$\begin{aligned} &\phi(x)F_1(x) + \Big(1-\phi(x)\Big)F_2(x) \geq \\ \geq\ & \phi(x) \min\{F_1(x), F_2(x)\} + \Big(1-\phi(x)\Big) \min\{F_1(x), F_2(x)\} \\ =\ & \min\{F_1(x), F_2(x)\}. \end{aligned}$$

Por lo tanto, para cualquier valor de x:

$$\phi(x)F_1(x) + \Big(1-\phi(x)\Big)F_2(x) \geq \min\{F_1(x), F_2(x)\}. \tag{2.11}$$

Sabemos que $0 \leq \chi(x) \leq 1$ por lo cual

$$\chi(x)\left(\phi(x)F_1(x) + \big(1-\phi(x)\big)F_2(x)\right) \geq \chi(x)\big(\min\{F_1(x), F_2(x)\}\big). \tag{2.12}$$

$\therefore$ Sumando (2.12) sobre x se obtiene:

$$\sum_x \chi(x)\Big(\phi(x)F_1(x) + \big(1-\phi(x)\big)F_2(x)\Big) \geq \sum_x \chi(x)\big(\min\{F_1(x), F_2(x)\}\big).$$

Ahora supongamos que $\phi(x)$ es la función indicadora del evento $\{F_1 \leq F_2\}$ *i. e.* $\phi(x) = I_{\{F_1 \leq F_2\}}(x)$. Recordemos que el evento $\{F_1 \leq F_2\} = \{x : F_1(x) \leq F_2(x)\}$. Sea $A = \{x : F_1(x) \leq F_2(x)\}$ por lo cual $A^c = \{x : F_1(x) > F_2(x)\}$ entonces:

$$\begin{aligned}
&\sum_x \chi(x)\Big(\phi(x)F_1(x) + \big(1-\phi(x)\big)F_2(x)\Big) = \\
= &\sum_{x\in A} \chi(x)\Big(\phi(x)F_1(x) + \big(1-\phi(x)\big)F_2(x)\Big) + \sum_{x\in A^c} \chi(x)\Big(\phi(x)F_1(x) + \big(1-\phi(x)\big)F_2(x)\Big)
\end{aligned} \tag{2.13}$$

Sin embargo, cuando $F_1(x) \leq F_2(x)$ entonces $\phi(x) = 1$; y cuando $F_2(x) < F_1(x)$ entonces $\phi(x) = 0$ y por tanto (2.13) se convierte en:

$$\begin{aligned}
\sum_x \chi(x)\Big(\phi(x)F_1(x) + \big(1-\phi(x)\big)F_2(x)\Big) &= \sum_{x\in A} \chi(x)F_1(x) + \sum_{x\in A^c} \chi(x)F_2(x) \\
&= \sum_x \chi(x)\min\{F_1(x), F_2(x)\} \qquad (2.14)
\end{aligned}$$

Por lo tanto, para el caso en que $\phi(x) = I_{\{F_1<F_2\}}(x)$ la desigualdad (2.10) es en realidad una igualdad (2.14). ■

El siguiente teorema nos muestra que la función de Lagrange sobre la regla de decisión elegida ϕ^*, cuyas componentes tienen la estructura dada en (2.9), es menor comparada con otra función de Lagrange bajo cualquier otra regla de decisión ϕ.

Teorema 2.

Para cualquier regla de paro ψ *y cualquier regla de decisión* ϕ *tales que*

$$L(\psi,\phi) \geq L(\psi,\phi^*). \tag{2.15}$$

donde las componentes de ϕ^* *tienen la estructura dada en* (2.9).

Más aún, para el valor de $L(\psi,\phi^*)$ *se tiene la siguiente expresión*

$$L(\psi,\phi^*) = K(\psi) + \sum_{k=1}^{\infty} \sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} s_{\boldsymbol{n}}^{\psi} \min\{\lambda_0 f_0^{\boldsymbol{n}}, \lambda_1 f_1^{\boldsymbol{n}}\}. \tag{2.16}$$

donde $s_{\boldsymbol{n}}^{\psi}$ *está dada por la fórmula* (1.4) *y*

$$x^{\boldsymbol{n}} = \left(x_1^{(n_1)}, \ldots, x_k^{(n_k)}\right), \quad \boldsymbol{n} \in \mathcal{G}^k \tag{2.17}$$

Demostración.

Por la definición de la función de Lagrange (2.2) se demostrará que

$$\lambda_0\alpha(\psi,\phi) + \lambda_1\beta(\psi,\phi) \geq \lambda_0\alpha(\psi,\phi^*) + \lambda_1\beta(\psi,\phi^*), \tag{2.18}$$

y que

$$\lambda_0\alpha(\psi,\phi^*) + \lambda_1\beta(\psi,\phi^*) = \sum_{k=1}^{\infty} \sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} s_{\boldsymbol{n}}^{\psi} \min\{\lambda_0 f_0^{\boldsymbol{n}}, \lambda_1 f_1^{\boldsymbol{n}}\}. \tag{2.19}$$

La clave de esta demostración es el Lema anterior. Tomemos las definiciones $\alpha(\psi,\phi)$ y

$\beta(\psi,\phi)$. Entonces tomando el lado izquierdo de (2.18) y desarrollándola se tiene:

$$\begin{aligned}
&\lambda_0\alpha(\psi,\phi)+\lambda_1\beta(\psi,\phi) = \\
= \ &\lambda_0\sum_{k=1}^{\infty}\sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}} E_{\theta_0}\phi_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi} + \lambda_1\sum_{k=1}^{\infty}\sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}} E_{\theta_1}\big(1-\phi_{\boldsymbol{n}}\big) s_{\boldsymbol{n}}^{\psi} \\
= \ &\sum_{k=1}^{\infty}\sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}}\Big(\lambda_0 E_{\theta_0}\phi_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi} + \lambda_1 E_{\theta_1}\big(1-\phi_{\boldsymbol{n}}\big) s_{\boldsymbol{n}}^{\psi}\Big).
\end{aligned} \tag{2.20}$$

Ahora apliquemos la definición del valor esperado a las respectivas funciones indicadoras:

$$\begin{aligned}
&\lambda_0\alpha(\psi,\phi)+\lambda_1\beta(\psi,\phi) = \\
= \ &\sum_{k=1}^{\infty}\sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}}\left(\lambda_0\sum_{x^{\boldsymbol{n}}}\phi_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi} f_0^{\boldsymbol{n}} + \lambda_1\sum_{x^{\boldsymbol{n}}}\big(1-\phi_{\boldsymbol{n}}\big) s_{\boldsymbol{n}}^{\psi} f_1^{\boldsymbol{n}}\right) \\
= \ &\sum_{k=1}^{\infty}\sum_{\boldsymbol{n}\in\mathcal{G}^k}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}}\Big(\phi_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi}\lambda_0 f_0^{\boldsymbol{n}} + \big(1-\phi_{\boldsymbol{n}}\big) s_{\boldsymbol{n}}^{\psi}\lambda_1 f_1^{\boldsymbol{n}}\Big) \\
= \ &\sum_{k=1}^{\infty}\sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}}\sum_{x^{\boldsymbol{n}}} s_{\boldsymbol{n}}^{\psi}\Big(\phi_{\boldsymbol{n}}\lambda_0 f_0^{\boldsymbol{n}} + \big(1-\phi_{\boldsymbol{n}}\big)\lambda_1 f_1^{\boldsymbol{n}}\Big).
\end{aligned} \tag{2.21}$$

Aplicando el Lema 1 a $s_{\boldsymbol{n}}^{\psi}\left(\phi_{\boldsymbol{n}}\lambda_0 f_0^{\boldsymbol{n}} + \big(1-\phi_{\boldsymbol{n}}\big)\lambda_1 f_1^{\boldsymbol{n}}\right)$ en el lado derecho de (2.21) donde

$$s_{\boldsymbol{n}}^{\psi}\equiv\chi(x) \qquad \phi_{\boldsymbol{n}}\equiv\phi(x) \qquad \text{y} \qquad \lambda_0 f_0^{\boldsymbol{n}}\equiv F_1(x), \qquad \lambda_1 f_1^{\boldsymbol{n}}\equiv F_2(x).$$

Entonces

$$\begin{aligned}
&\sum_{k=1}^{\infty}\sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}}\sum_{x^{\boldsymbol{n}}} s_{\boldsymbol{n}}^{\psi}\Big(\phi_{\boldsymbol{n}}\lambda_0 f_0^{\boldsymbol{n}} + \big(1-\phi_{\boldsymbol{n}}\big)\lambda_1 f_1^{\boldsymbol{n}}\Big) \\
&\qquad \geq \ \sum_{k=1}^{\infty}\sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}}\sum_{x^{\boldsymbol{n}}} s_{\boldsymbol{n}}^{\psi}\,\text{mín}\,\{\lambda_0 f_0^{\boldsymbol{n}},\lambda_1 f_1^{\boldsymbol{n}}\}.
\end{aligned} \tag{2.22}$$

Por lo tanto

$$\lambda_0\alpha(\psi,\phi)+\lambda_1\beta(\psi,\phi) \ \geq \ \sum_{k=1}^{\infty}\sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}}\sum_{x^{\boldsymbol{n}}} s_{\boldsymbol{n}}^{\psi}\,\text{mín}\,\{\lambda_0 f_0^{\boldsymbol{n}},\lambda_1 f_1^{\boldsymbol{n}}\}.$$

También sabemos por el Lema 1 que si $\phi_{\boldsymbol{n}} = I_{\{\lambda_0 f_0^{\boldsymbol{n}} \leq \lambda_1 f_1^{\boldsymbol{n}}\}}$ entonces se obtiene la igualdad, por tanto tomemos $\phi_{\boldsymbol{n}} = \phi_{\boldsymbol{n}}^*$ entonces

$$\lambda_0 \alpha(\psi, \phi^*) + \lambda_1 \beta(\psi, \phi^*) = \sum_{k=1}^{\infty} \sum_{\boldsymbol{n} \in \mathcal{G}^k} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} s_{\boldsymbol{n}}^{\psi} \min \{\lambda_0 f_0^{\boldsymbol{n}}, \lambda_1 f_1^{\boldsymbol{n}}\}.$$

Estas últimas relaciones demuestran (2.18) y (2.19) y por tanto demuestran las relaciones (2.15) y (2.16) ■

El teorema 2 demuestra que la función de Lagrange $L(\psi, \phi^*)$ con la regla de decisión ϕ^* fija es la menor sobre cualquier otra función de Lagrange $L(\psi, \phi)$ para cualquier otra regla de decisión ϕ. De esta manera el problema de minimización de $L(\psi, \phi)$ se convierte ahora en un problema de paro óptimo, ya que tomaremos $L(\psi)$ que dependerá únicamente de la regla de paro, mediante la siguiente asignación.

Notación: Sea

$$L(\psi, \phi^*) = L(\psi). \tag{2.23}$$

Entonces, el objetivo, a partir de este momento, será hallar la regla de paro ψ^* tal que

$$L(\psi) \geq L(\psi^*). \tag{2.24}$$

para cualquier regla de paro ψ, *i. e.*

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

Con esto será resuelto el problema de la prueba óptima ya que para cualquier prueba (ψ, ϕ), por el Teorema 2 se obtendrá:

$$L(\psi, \phi) \geq L(\psi, \phi^*) = L(\psi) \geq L(\psi^*) = L(\psi^*, \phi^*) \tag{2.25}$$

lo cual es lo que se quería. Recuérdese que el teorema 1 garantiza que la prueba optimizará la función de Lagrange $L(\psi,\phi) \geq L(\psi^*,\phi^*)$ (2.3) sobre cualquier otra prueba. Así la prueba (ψ^*,ϕ^*) será la óptima.

Sabemos por teorema 2 que

$$L(\psi,\phi^*) = K(\psi) + \sum_{k=1}^{\infty}\sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} s_{\boldsymbol{n}}^{\psi} \operatorname{mín}\{\lambda_0 f_0^{\boldsymbol{n}}, \lambda_1 f_1^{\boldsymbol{n}}\}.$$

Reescribamos esta última relación sustituyendo $K(\psi)$ dada por (1.27).

Notación: Sea

$$l_{\boldsymbol{n}} = \operatorname{mín}\left\{\lambda_0 f_0^{\boldsymbol{n}}, \lambda_1 f_1^{\boldsymbol{n}}\right\} \tag{2.26}$$

Entonces para $\psi \in \Psi_{\theta_0}$,

$$\begin{aligned} L(\psi) &= \sum_{k=1}^{\infty}\sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}} c(\boldsymbol{n})\; E_{\theta_0} s_{\boldsymbol{n}}^{\psi} + \sum_{k=1}^{\infty}\sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} s_{\boldsymbol{n}}^{\psi} l_{\boldsymbol{n}}. \\ &= \sum_{k=1}^{\infty}\sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} c(\boldsymbol{n}) s_{\boldsymbol{n}}^{\psi} f_0^{\boldsymbol{n}} + \sum_{k=1}^{\infty}\sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} s_{\boldsymbol{n}}^{\psi} l_{\boldsymbol{n}} \qquad (2.27) \\ &= \sum_{k=1}^{\infty}\sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} s_{\boldsymbol{n}}^{\psi} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big). \qquad (2.28) \end{aligned}$$

Teniendo en cuenta que $E_{\theta_0} s_{\boldsymbol{n}}^{\psi} = \sum_{x^{\boldsymbol{n}}} s_{\boldsymbol{n}}^{\psi} f_0^{\boldsymbol{n}}$.

2.2.2. Paro óptimo. Caso truncado

En esta sección vamos a resolver el problema, planteado en la sección anterior, sobre el paro óptimo, pero de manera "parcial", al cual llamaremos el *caso truncado*.

En el contexto de un experimento con su análisis secuencial, puede suceder que las respectivas reglas de paro, conforme vamos analizando los datos de los grupos, sean siempre iguales a cero $\left(\psi_{\boldsymbol{n}}\left(x^{(n_1)},\ldots,x^{(n_k)}\right)=0, \boldsymbol{n}\in\mathcal{G}^k, k=1,2,\ldots\right)$, *i. e.* puede caber la posibilidad de que no se vea el termino del experimento estadístico. Para casos como éste debemos tener en cuenta una condición de paro que trunque (o detenga el análisis) para que el experimento no continue indefinidamente, ya que esto no sería posible. A este caso lo designamos el *caso truncado*, ya que definiremos un número **máximo** de etapas de paro en el experimento. Cuando se llegue a este número, el experimento se terminará. Otra manera de ver el caso truncado es fijar de antemano un número máximo de grupos observados independientemente de los valores anteriores de las reglas de paro. Es decir, también se puede tomar el número fijo y realizar el análisis simplemente hasta llegar a él dejando variable la información previa.

Sea N la etapa de paro máxima en un experimento con grupos de tamaños aleatorio representados en el vector $\boldsymbol{n}=(n_1,n_2,\ldots,n_N)\in\mathcal{G}^N$. Entonces, el análisis secuencial aplicará las correspondientes componentes de la regla de paro a cada grupo de observaciones hasta el N-ésimo grupo para el cual $\psi_{\boldsymbol{n}}\equiv 1$ con $\boldsymbol{n}\in\mathcal{G}^N$. Ello significa que se aplicarán las componentes de la regla de paro a $(N-1)$ grupos $\psi_{(n:1)},\psi_{(n:2)},\ldots,\psi_{(n:N-1)}$ y la aplicación de la N-ésima componente terminará el experimento.

Ahora le daremos a la función $L(\psi)$, para el caso truncado con $\boldsymbol{n}\in\mathcal{G}^k$, $(k=1,\ldots,N)$, una forma más conveniente utilizando la relación del costo total $K(\theta_0;\psi)$.

Sea $L_N(\psi)$ la función de *Lagrange Truncada* definida como:

$$\begin{aligned} L_N(\psi) &= K_N(\psi) + \sum_{k=1}^{N} \sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} s_{\boldsymbol{n}}^{\psi} l_{\boldsymbol{n}} \\ &= \sum_{k=1}^{N} \sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} c(\boldsymbol{n}) s_{\boldsymbol{n}}^{\psi} f_0^{\boldsymbol{n}} + \sum_{k=1}^{N} \sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} s_{\boldsymbol{n}}^{\psi} l_{\boldsymbol{n}}. \\ &= \sum_{k=1}^{N} \sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} s_{\boldsymbol{n}}^{\psi} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big). \end{aligned} \tag{2.29}$$

Tengamos en cuenta que $\psi_{\boldsymbol{n}} \equiv 1$ con $\boldsymbol{n} \in \mathcal{G}^N$ y recordemos las definiciones de $t_{\boldsymbol{n}}^{\psi}$ (1.2) y de $s_{\boldsymbol{n}}^{\psi}$ (1.4). Entonces (2.29) se convierte en:

$$L_N(\psi) = \sum_{k=1}^{N-1} \sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} s_{\boldsymbol{n}}^{\psi} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big) + \sum_{\boldsymbol{n}\in\mathcal{G}^N} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} t_{\boldsymbol{n}}^{\psi} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big) \tag{2.30}$$

El siguiente lema absorbe la mayor parte de la carga técnica en el desarrollo de la regla truncada.

Lema 2.

Sea $r \geq 2$ *fijo y sea* $v_{\boldsymbol{n}} = v_{\boldsymbol{n}}\Big(x_1^{(n_1)}, x_2^{(n_2)}, \ldots, x_r^{(n_r)}\Big)$ *con* $\boldsymbol{n} \in \mathcal{G}^r$ *una familia de funciones cualquiera. Entonces*

$$\sum_{k=1}^{r-1} \sum_{\boldsymbol{n}\in\mathcal{G}^k} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}}\Big) + \sum_{\boldsymbol{n}\in\mathcal{G}^r} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + v_{\boldsymbol{n}}\Big) \geq$$
$$\geq \sum_{k=1}^{r-2} \sum_{\boldsymbol{n}\in\mathcal{G}^k} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}}\Big) + \sum_{\boldsymbol{n}\in\mathcal{G}^{r-1}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + w_{\boldsymbol{n}}\Big) \tag{2.31}$$

en donde para cualquier $\boldsymbol{n} \in \mathcal{G}^{r-1}$,

$$w_{\boldsymbol{n}} = \text{mín}\Big\{l_{\boldsymbol{n}}, \bar{c} f_0^{\boldsymbol{n}} + \sum_{n_r\in\mathcal{G}} q_{n_r} \sum_{x_r^{(n_r)}} v_{(\boldsymbol{n},n_r)}\Big\} \tag{2.32}$$

y la desigualdad en (2.31) *se convierte en igualdad si*

$$\psi_{\boldsymbol{n}} = I_{\Big\{l_{\boldsymbol{n}} \,\leq\, \bar{c} f_0^{\boldsymbol{n}} + \sum_{n_r\in\mathcal{G}} q_{n_r} \sum_{x_r^{(n_r)}} v_{(\boldsymbol{n},n_r)}\Big\}} \tag{2.33}$$

para cualquier $\boldsymbol{n} \in \mathcal{G}^{r-1}$.

Demostración. Tomemos el lado izquierdo de la ecuación (2.31) y desarrollemoslo.

$$\sum_{k=1}^{r-1} \sum_{\boldsymbol{n}\in\mathcal{G}^k} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}}\Big) + \sum_{\boldsymbol{n}\in\mathcal{G}^r} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + v_{\boldsymbol{n}}\Big) =$$
$$= \sum_{k=1}^{r-2} \sum_{\boldsymbol{n}\in\mathcal{G}^k} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}}\Big) +$$
$$+ \sum_{\boldsymbol{n}\in\mathcal{G}^{r-1}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}}\Big) + \sum_{\boldsymbol{n}\in\mathcal{G}^r} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + v_{\boldsymbol{n}}\Big) \tag{2.34}$$

Ahora sólo desarrollemos los últimos dos sumandos de la ecuación anterior (2.34).

$$\sum_{\boldsymbol{n}\in\mathcal{G}^{r-1}}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}}+l_{\boldsymbol{n}}\Big)+\sum_{\boldsymbol{n}\in\mathcal{G}^{r}}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}}+v_{\boldsymbol{n}}\Big)=$$

$$=\sum_{\boldsymbol{n}\in\mathcal{G}^{r-1}}\left(\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi}\big(c(\boldsymbol{n})f_0^{\boldsymbol{n}}+l_{\boldsymbol{n}}\big)+\right.$$

$$\left.+\sum_{n_r\in\mathcal{G}}\sum_{\left(x_1^{(n_1)},x_2^{(n_2)},\dots,x_{r-1}^{(n_{r-1})},x_r^{(n_r)}\right)} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi}\Big(\big(c(\boldsymbol{n})+c(n_r)\big)f_0^{(\boldsymbol{n},n_r)}+v_{(\boldsymbol{n},n_r)}\Big)\right)$$

$$=\sum_{\boldsymbol{n}\in\mathcal{G}^{r-1}}\left(\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi}\big(c(\boldsymbol{n})f_0^{\boldsymbol{n}}+l_{\boldsymbol{n}}\big)+\sum_{x^{\boldsymbol{n}}}\sum_{n_r\in\mathcal{G}}\sum_{x_r^{(n_r)}} q_{n_1}q_{n_2},\dots,q_{n_{r-1}}q_{n_r}\right.$$

$$\Big((1-\psi_{(\boldsymbol{n}:1)})(1-\psi_{(\boldsymbol{n}:2)})\dots(1-\psi_{(\boldsymbol{n}:r-2)})(1-\psi_{(\boldsymbol{n}:r-1)})\Big)$$

$$\left.\Big(\big(c(\boldsymbol{n})+c(n_r)\big)f_0^{(\boldsymbol{n},n_r)}+v_{(\boldsymbol{n},n_r)}\Big)\right)$$

$$=\sum_{\boldsymbol{n}\in\mathcal{G}^{r-1}}\left(\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}}(1-\psi_{(\boldsymbol{n}:1)})(1-\psi_{(\boldsymbol{n}:2)})\dots(1-\psi_{(\boldsymbol{n}:r-2)})\psi_{(\boldsymbol{n}:r-1)}\big(c(\boldsymbol{n})f_0^{\boldsymbol{n}}+l_{\boldsymbol{n}}\big)+\right.$$

$$+\sum_{x^{\boldsymbol{n}}}\sum_{n_r\in\mathcal{G}}\sum_{x_r^{(n_r)}} q_{\boldsymbol{n}}q_{n_r}\Big((1-\psi_{(\boldsymbol{n}:1)})(1-\psi_{(\boldsymbol{n}:2)})\dots(1-\psi_{(\boldsymbol{n}:r-2)})(1-\psi_{(\boldsymbol{n}:r-1)})\Big)$$

$$\left.\Big(\big(c(\boldsymbol{n})+c(n_r)\big)f_0^{(\boldsymbol{n},n_r)}+v_{(\boldsymbol{n},n_r)}\Big)\right)$$

$$=\sum_{\boldsymbol{n}\in\mathcal{G}^{r-1}}\left(\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi}\left[\psi_{\boldsymbol{n}}\big(c(\boldsymbol{n})f_0^{\boldsymbol{n}}+l_{\boldsymbol{n}}\big)+\right.\right.$$

$$\left.\left.+\sum_{n_r\in\mathcal{G}}\sum_{x_r^{(n_r)}} q_{n_r}\Big(1-\psi_{\boldsymbol{n}}\Big)\Big(\big(c(\boldsymbol{n})+c(n_r)\big)f_0^{(\boldsymbol{n},n_r)}+v_{(\boldsymbol{n},n_r)}\Big)\right]\right)$$

$$=\sum_{\boldsymbol{n}\in\mathcal{G}^{r-1}} q_{\boldsymbol{n}}\left(\sum_{x^{\boldsymbol{n}}} t_{\boldsymbol{n}}^{\psi}\left[\psi_{\boldsymbol{n}}\big(c(\boldsymbol{n})f_0^{\boldsymbol{n}}+l_{\boldsymbol{n}}\big)+\right.\right.$$

$$\left.\left.+\Big(1-\psi_{\boldsymbol{n}}\Big)\sum_{n_r\in\mathcal{G}} q_{n_r}\sum_{x_r^{(n_r)}}\Big(\big(c(\boldsymbol{n})+c(n_r)\big)f_0^{(\boldsymbol{n},n_r)}+v_{(\boldsymbol{n},n_r)}\Big)\right]\right)$$

$$= \sum_{\boldsymbol{n}\in\mathcal{G}^{r-1}} q_{\boldsymbol{n}} \Bigg(\sum_{x^{\boldsymbol{n}}} t^{\psi}_{\boldsymbol{n}} \Bigg[\psi_{\boldsymbol{n}} \big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \big) +$$

$$+ \Big(1 - \psi_{\boldsymbol{n}}\Big) \Big(c(\boldsymbol{n}) \sum_{n_r\in\mathcal{G}} q_{n_r} \sum_{x_r^{(n_r)}} f_0^{(\boldsymbol{n},n_r)} + \sum_{n_r\in\mathcal{G}} q_{n_r} \sum_{x_r^{(n_r)}} \big(c(n_r) f_0^{(\boldsymbol{n},n_r)} + v_{(\boldsymbol{n},n_r)} \big) \Big) \Bigg] \Bigg)$$

Ahora tengamos en cuenta que $\sum\limits_{x_r^{(n_r)}} f_0^{(\boldsymbol{n},n_r)} = f_0^{\boldsymbol{n}}$ debido a que se tiene una suma marginal sobre el vector $x_r^{(n_r)}$. Tomando en cuenta este último hecho, recordemos que $\sum\limits_{\boldsymbol{n}\in\mathcal{G}} q_{\boldsymbol{n}} = 1$, entonces la relación anterior se puede escribir como:

$$\sum_{\boldsymbol{n}\in\mathcal{G}^{r-1}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s^{\psi}_{\boldsymbol{n}} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big) + \sum_{\boldsymbol{n}\in\mathcal{G}^{r}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t^{\psi}_{\boldsymbol{n}} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + v_{\boldsymbol{n}} \Big) =$$

$$= \sum_{\boldsymbol{n}\in\mathcal{G}^{r-1}} q_{\boldsymbol{n}} \Bigg[\sum_{x^{\boldsymbol{n}}} t^{\psi}_{\boldsymbol{n}} \Bigg(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + \psi_{\boldsymbol{n}} l_{\boldsymbol{n}} +$$

$$+ \Big(1 - \psi_{\boldsymbol{n}}\Big) \Big(\sum_{n_r\in\mathcal{G}} q_{n_r} \sum_{x_r^{(n_r)}} \big(c(n_r) f_0^{(\boldsymbol{n},n_r)} + v_{(\boldsymbol{n},n_r)} \big) \Big) \Bigg) \Bigg]$$

$$= \sum_{\boldsymbol{n}\in\mathcal{G}^{r-1}} q_{\boldsymbol{n}} \Bigg[\sum_{x^{\boldsymbol{n}}} t^{\psi}_{\boldsymbol{n}} \Bigg(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + \psi_{\boldsymbol{n}} l_{\boldsymbol{n}} +$$

$$+ \Big(1 - \psi_{\boldsymbol{n}}\Big) \Big(f_0^{\boldsymbol{n}} \sum_{n_r\in\mathcal{G}} q_{n_r} c(n_r) + \sum_{n_r\in\mathcal{G}} q_{n_r} \sum_{x_r^{(n_r)}} v_{(\boldsymbol{n},n_r)} \Big) \Bigg) \Bigg]$$

Pero $c(n_r) = c_{n_r}$, por lo cual $\sum\limits_{n_r\in\mathcal{G}} q_{n_r} c_{n_r} = \overline{c}$.

$$\begin{aligned}
\therefore \qquad & \sum_{\boldsymbol{n}\in\mathcal{G}^{r-1}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}}\Big) + \sum_{\boldsymbol{n}\in\mathcal{G}^{r}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + v_{\boldsymbol{n}}\Big) = \\
= \quad & \sum_{\boldsymbol{n}\in\mathcal{G}^{r-1}} q_{\boldsymbol{n}} \Bigg[\sum_{x^{\boldsymbol{n}}} t_{\boldsymbol{n}}^{\psi} \Bigg(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + \psi_{\boldsymbol{n}} l_{\boldsymbol{n}} + \\
& \qquad + \Big(1 - \psi_{\boldsymbol{n}}\Big)\Big(\bar{c} f_0^{\boldsymbol{n}} + \sum_{n_r\in\mathcal{G}} q_{n_r} \sum_{x_r^{(n_r)}} v_{(\boldsymbol{n},n_r)}\Big)\Bigg)\Bigg]
\end{aligned} \tag{2.35}$$

Ahora sustituyamos la ecuación (2.35) en la parte correspondiente de (2.34) y se obtiene:

$$\begin{aligned}
& \sum_{k=1}^{r-2} \sum_{\boldsymbol{n}\in\mathcal{G}^{k}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}}\Big) + \\
& + \sum_{\boldsymbol{n}\in\mathcal{G}^{r-1}} q_{\boldsymbol{n}} \Bigg[\sum_{x^{\boldsymbol{n}}} t_{\boldsymbol{n}}^{\psi} \Bigg(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + \psi_{\boldsymbol{n}} l_{\boldsymbol{n}} + \Big(1 - \psi_{\boldsymbol{n}}\Big)\Big(\bar{c} f_0^{\boldsymbol{n}} + \sum_{n_r\in\mathcal{G}} q_{n_r} \sum_{x_r^{(n_r)}} v_{(\boldsymbol{n},n_r)}\Big)\Bigg)\Bigg]
\end{aligned}$$

Apliquemos el lema 1 (2.10) a esta última relación, donde:

$$\chi(x) \equiv t_{\boldsymbol{n}}^{\psi}, \quad \phi(x) \equiv \psi_{\boldsymbol{n}}, \quad F_1 \equiv l_{\boldsymbol{n}} \quad \text{y} \quad F_2 \equiv \bar{c} f_0^{\boldsymbol{n}} + \sum_{n_r\in\mathcal{G}} q_{n_r} \sum_{x_r^{(n_r)}} v_{(\boldsymbol{n},n_r)}$$

Por tanto:

$$\begin{aligned}
& \sum_{k=1}^{r-2} \sum_{\boldsymbol{n}\in\mathcal{G}^{k}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}}\Big) + \\
& \quad + \sum_{\boldsymbol{n}\in\mathcal{G}^{r-1}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi} \Bigg(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + \psi_{\boldsymbol{n}} l_{\boldsymbol{n}} + \Big(1 - \psi_{\boldsymbol{n}}\Big)\Big(\bar{c} f_0^{\boldsymbol{n}} + \sum_{n_r\in\mathcal{G}} q_{n_r} \sum_{x_r^{(n_r)}} v_{(\boldsymbol{n},n_r)}\Big)\Bigg) \\
\geq \quad & \sum_{k=1}^{r-2} \sum_{\boldsymbol{n}\in\mathcal{G}^{k}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}}\Big) + \\
& \quad + \sum_{\boldsymbol{n}\in\mathcal{G}^{r-1}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi} \Bigg(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + \text{mín}\Big\{ l_{\boldsymbol{n}}, \bar{c} f_0^{n} + \sum_{n_r\in\mathcal{G}} q_{n_r} \sum_{x_r^{(n_r)}} v_{(\boldsymbol{n},n_r)} \Big\} \Bigg)
\end{aligned} \tag{2.36}$$

Además, recordemos que el lema 1 nos indica cuándo la desigualdad (2.10) se convierte en igualdad y esto es cuando $\big(\phi(x) = I_{\{F_1 \leq F_2\}}(x)\big)$ y en nuestro caso cuando para cualquier $\boldsymbol{n} \in \mathcal{G}^{r-1}$,

$$\psi_{\boldsymbol{n}} = I_{\left\{l_{\boldsymbol{n}} \ \leq \ \bar{c} f_0^{\boldsymbol{n}} + \sum\limits_{n_r \in \mathcal{G}} q_{n_r} \sum\limits_{x_r^{(n_r)}} v_{(\boldsymbol{n}, n_r)}\right\}}. \tag{2.37}$$

Juntemos ahora la ecuaciones (2.34) y (2.36) obteniendo:

$$\begin{aligned}
&\sum_{k=1}^{r-1} \sum_{\boldsymbol{n} \in \mathcal{G}^k} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big) + \sum_{\boldsymbol{n} \in \mathcal{G}^r} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + v_{\boldsymbol{n}} \Big) \geq \\
&\quad \geq \sum_{k=1}^{r-2} \sum_{\boldsymbol{n} \in \mathcal{G}^k} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big) + \\
&\qquad + \sum_{\boldsymbol{n} \in \mathcal{G}^{r-1}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi} \left(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + \text{mín} \left\{ l_{\boldsymbol{n}}, \bar{c} f_0^{\boldsymbol{n}} + \sum_{n_r \in \mathcal{G}} q_{n_r} \sum_{x_r^{(n_r)}} v_{(\boldsymbol{n}, n_r)} \right\} \right)
\end{aligned} \tag{2.38}$$

Lo cual demuestra el lema 2 y la igualdad se da cuando sucede (2.37). ■

Ahora aplicaremos el lema 2 (2.31) a la ecuación de Lagrange truncada (2.30), obteniendo:

$$\begin{aligned}
L_N(\psi) &= \sum_{k=1}^{N-1} \sum_{\boldsymbol{n} \in \mathcal{G}^k} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big) + \sum_{\boldsymbol{n} \in \mathcal{G}^N} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big) \\
&\geq \sum_{k=1}^{N-2} \sum_{\boldsymbol{n} \in \mathcal{G}^k} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big) + \sum_{\boldsymbol{n} \in \mathcal{G}^{N-1}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + w_{\boldsymbol{n}} \Big)
\end{aligned}$$

en donde para $\boldsymbol{n} \in \mathcal{G}^{N-1}$,

$$w_{\boldsymbol{n}} = \text{mín} \left\{ l_{\boldsymbol{n}}, \bar{c} f_0^{\boldsymbol{n}} + \sum_{n_N \in \mathcal{G}} q_{n_N} \sum_{x_N^{(n_N)}} v_{(\boldsymbol{n}, n_N)} \right\}$$

Apliquemos, nuevamente, al último termino de la desigualdad anterior el lema 2 y así sucesivamente apliquese $N-1$ veces obteniendo la siguiente cadena:

$$
\begin{aligned}
L_N(\psi) &= \sum_{k=1}^{N-1}\sum_{\boldsymbol{n}\in\mathcal{G}^k}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}}+l_{\boldsymbol{n}}\Big)+\sum_{\boldsymbol{n}\in\mathcal{G}^N}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}}+l_{\boldsymbol{n}}\Big) \\
&\geq \sum_{k=1}^{N-2}\sum_{\boldsymbol{n}\in\mathcal{G}^k}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}}+l_{\boldsymbol{n}}\Big)+\sum_{\boldsymbol{n}\in\mathcal{G}^{N-1}}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}}+w_{\boldsymbol{n}}\Big) \qquad \text{paso } 1 \\
&\geq \sum_{k=1}^{N-3}\sum_{\boldsymbol{n}\in\mathcal{G}^k}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}}+l_{\boldsymbol{n}}\Big)+\sum_{\boldsymbol{n}\in\mathcal{G}^{N-2}}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}}+w_{\boldsymbol{n}}\Big) \qquad \text{paso } 2 \\
&\vdots \\
&\geq \sum_{k=1}^{2}\sum_{\boldsymbol{n}\in\mathcal{G}^k}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}}+l_{\boldsymbol{n}}\Big)+\sum_{\boldsymbol{n}\in\mathcal{G}^{3}}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}}+w_{\boldsymbol{n}}\Big) \qquad \text{paso } N-3 \\
&\geq \sum_{k=1}^{1}\sum_{\boldsymbol{n}\in\mathcal{G}^k}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}}+l_{\boldsymbol{n}}\Big)+\sum_{\boldsymbol{n}\in\mathcal{G}^{2}}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}}+w_{\boldsymbol{n}}\Big) \qquad \text{paso } N-2 \\
&\geq \sum_{\boldsymbol{n}\in\mathcal{G}}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}}+w_{\boldsymbol{n}}\Big) \qquad \text{paso } N-1
\end{aligned}
\tag{2.39}
$$

Donde para cada uno de los pasos se tienen sus correspondientes funciones.

Paso 1: Para cualquier $\boldsymbol{n}\in\mathcal{G}^{N-1}$

$$
w_{\boldsymbol{n}} = \min\Big\{l_{\boldsymbol{n}}, \bar{c}f_0^{\boldsymbol{n}}+\sum_{n_N\in\mathcal{G}} q_{n_N}\sum_{x_N^{(n_N)}} l_{(\boldsymbol{n},n_N)}\Big\} = V_{\boldsymbol{n}}^N \tag{2.40}
$$

y la igualdad se alcanza cuando

$$
\psi_{\boldsymbol{n}} = I_{\Big\{l_{\boldsymbol{n}} \leq \bar{c}f_0^{\boldsymbol{n}}+\sum_{n_N\in\mathcal{G}} q_{n_N}\sum_{x_N^{(n_N)}} l_{(\boldsymbol{n},n_N)}\Big\}}. \tag{2.41}
$$

Observación: En la función $V_{\boldsymbol{n}}^N$ el vector $\boldsymbol{n}$ será diferente para cada paso, éste tendrá dimensión menor conforme se avance. Sin embargo el superíndice N será el mismo para todos los pasos, N nos indica el inicio de la dimensión de la cadena en cuestión.

Paso 2: Para $\boldsymbol{n} \in \mathcal{G}^{N-2}$,

$$w_{\boldsymbol{n}} = \text{mín}\left\{l_{\boldsymbol{n}}, \bar{c}f_0^{\boldsymbol{n}} + \sum_{n_{N-1}\in\mathcal{G}} q_{n_{N-1}} \sum_{x_{N-1}^{(n_{N-1})}} w_{(\boldsymbol{n},n_{N-1})}\right\} = V_{\boldsymbol{n}}^N \tag{2.42}$$

y la igualdad se alcanza, cuando

$$\psi_{\boldsymbol{n}} = I_{\left\{l_{\boldsymbol{n}} \,\leq\, \bar{c}f_0^{\boldsymbol{n}} + \sum_{n_{N-1}\in\mathcal{G}} q_{n_{N-1}} \sum_{x_{N-1}^{(n_{N-1})}} w_{(\boldsymbol{n},n_{N-1})}\right\}} \tag{2.43}$$

$\vdots$

Paso $\boldsymbol{N-2}$: Con $\boldsymbol{n} \in \mathcal{G}^2$

$$w_{\boldsymbol{n}} = \text{mín}\left\{l_{\boldsymbol{n}}, \bar{c}f_0^{\boldsymbol{n}} + \sum_{n_3\in\mathcal{G}} q_{n_3} \sum_{x_3^{(n_3)}} w_{(\boldsymbol{n},n_3)}\right\} = V_{\boldsymbol{n}}^N$$

y la igualdad se alcanza, cuando $\forall\ n_2 \in \mathcal{G}$

$$\psi_{\boldsymbol{n}} = I_{\left\{l_{\boldsymbol{n}} \,\leq\, \bar{c}f_0^{\boldsymbol{n}} + \sum_{n_3\in\mathcal{G}} q_{n_3} \sum_{x_3^{(n_3)}} w_{(\boldsymbol{n},n_3)}\right\}}$$

Paso $\boldsymbol{N-1}$: Con $\boldsymbol{n} \in \mathcal{G}$

$$w_{\boldsymbol{n}} = \text{mín}\left\{l_{\boldsymbol{n}}, \bar{c}f_0^{\boldsymbol{n}} + \sum_{n_2\in\mathcal{G}} q_{n_2} \sum_{x_2^{(n_2)}} w_{(\boldsymbol{n},n_2)}\right\} = V_{\boldsymbol{n}}^N \tag{2.44}$$

y la igualdad, cuando

$$\psi_{\boldsymbol{n}} = I_{\left\{l_{\boldsymbol{n}} \,\leq\, \bar{c}f_0^{\boldsymbol{n}} + \sum_{n_2\in\mathcal{G}} q_{n_2} \sum_{x_2^{(n_2)}} w_{(\boldsymbol{n},n_2)}\right\}}. \tag{2.45}$$

Ahora desarrollemos el último término de la cadena anterior de desigualdades (2.39), esto es desarrollemos el término del paso $N-1$. Se sabe por la definición (1.2) que cuando $\boldsymbol{n} \in \mathcal{G}$ entonces $t_{\boldsymbol{n}}^{\psi} = 1$, por lo cual

$$\begin{aligned}
\sum_{\boldsymbol{n}\in\mathcal{G}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + w_{\boldsymbol{n}} \Big) &= \sum_{\boldsymbol{n}\in\mathcal{G}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + w_{\boldsymbol{n}} \Big) \\
&= \sum_{\boldsymbol{n}\in\mathcal{G}} q_{\boldsymbol{n}} c(\boldsymbol{n}) \sum_{x^{\boldsymbol{n}}} f_0^{\boldsymbol{n}} + \sum_{\boldsymbol{n}\in\mathcal{G}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} w_{\boldsymbol{n}} \\
&= \sum_{\boldsymbol{n}\in\mathcal{G}} q_{\boldsymbol{n}} c(\boldsymbol{n}) + \sum_{\boldsymbol{n}\in\mathcal{G}} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} w_{\boldsymbol{n}} \\
&= \bar{c} + \sum_{\boldsymbol{n}\in\mathcal{G}} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} w_{\boldsymbol{n}} \qquad (2.46)
\end{aligned}$$

Por tanto tenemos una cota inferior para nuestra función de Lagrange truncada, a saber:

$$L_N(\psi) \geq \bar{c} + \sum_{\boldsymbol{n}\in\mathcal{G}} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} w_{\boldsymbol{n}} \qquad (2.47)$$

donde $w_{\boldsymbol{n}}$ está dada por (2.44) y la desigualdad se alcanza cuando se tiene la regla $\psi_{\boldsymbol{n}}$, sucesivamente para $N-1 \geq k \geq 1$ con $\boldsymbol{n} \in \mathcal{G}^k$. Lo cual significa que se tienen las condiciones suficientes para alcanzar la mínima cota inferior, la cual nos describirá la regla de paro óptima.

Toda esta argumentación se formaliza en el siguiente lema:

Lema 3.

Sea ψ cualquier regla de paro truncada $\big(\psi_{\boldsymbol{n}} \equiv 1,\ \boldsymbol{n} \in \mathcal{G}^N\big)$. Entonces para cualquier $1 \leq r \leq N-1$ se cumplen las siguientes desigualdades:

$$
\begin{aligned}
L_N(\psi) &\geq \sum_{k=1}^{r} \sum_{\boldsymbol{n}\in\mathcal{G}^k} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s^{\psi}_{\boldsymbol{n}} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big) + \sum_{\boldsymbol{n}\in\mathcal{G}^{r+1}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t^{\psi}_{\boldsymbol{n}} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + V^N_{\boldsymbol{n}} \Big) \\
&\geq \sum_{k=1}^{r-1} \sum_{\boldsymbol{n}\in\mathcal{G}^k} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s^{\psi}_{\boldsymbol{n}} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big) + \sum_{\boldsymbol{n}\in\mathcal{G}^{r}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t^{\psi}_{\boldsymbol{n}} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + V^N_{\boldsymbol{n}} \Big)
\end{aligned}
\tag{2.48}
$$

donde $V^N_{\boldsymbol{n}} = l_{\boldsymbol{n}}$ con $\boldsymbol{n} \in \mathcal{G}^N$ y para $r < N$ y $\boldsymbol{n} \in \mathcal{G}^r$ recursivamente se definen:

$$
V^N_{\boldsymbol{n}} = \text{mín} \Big\{ l_{\boldsymbol{n}}, \bar{c} f_0^{\boldsymbol{n}} + \sum_{m\in\mathcal{G}} q_m \sum_{x_m^{(m)}} V^N_{(\boldsymbol{n},m)} \Big\}. \tag{2.49}
$$

Y $L_N(\psi)$ coincide con el lado derecho de (2.48) si $\forall\ k$ tal que $r \leq k \leq N-1$, $\boldsymbol{n} \in \mathcal{G}^k$

$$
\psi_{\boldsymbol{n}} = I_{\Big\{ l_{\boldsymbol{n}} \,\leq\, \bar{c} f_0^{\boldsymbol{n}} + \sum_{m\in\mathcal{G}} q_m \sum_{x_m^{(m)}} V^N_{(\boldsymbol{n},m)} \Big\}}. \tag{2.50}
$$

Demostración.

Sea $r = N-1$ entonces la desigualdad (2.48) se convierte exactamente en el **paso 1** de la cadena (2.39) con $V^N_{\boldsymbol{n}} = l_{\boldsymbol{n}}$, $\boldsymbol{n} \in \mathcal{G}^N$; y para $\boldsymbol{n} \in \mathcal{G}^{N-1}$, $V^N_{\boldsymbol{n}}$ y $\psi_{\boldsymbol{n}}$ están dadas, respectivamente, en (2.40) y (2.41). Cuando $r = N-2$ entonces la desigualdad (2.48) se convierte exactamente en el **paso 2** de la cadena (2.39) donde para $\boldsymbol{n} \in \mathcal{G}^{N-2}$: $V^N_{\boldsymbol{n}}$ y $\psi_{\boldsymbol{n}}$

están dadas,respectivamente, en (2.42) y (2.43). Así sucesivamente hasta que sea $r = 1$ entonces la desigualdad (2.48) se convierte exactamente en el **paso** $N - 1$ de la cadena (2.39) y para $\boldsymbol{n} \in \mathcal{G}$: $V_{\boldsymbol{n}}^N$ y $\psi_{\boldsymbol{n}}$ están dadas, respectivamente, en (2.44) y (2.45). ■

En términos del lema 3, la cota inferior (2.47) se puede reescribir como:

$$L_N(\psi) \quad \geq \quad \overline{c} + \sum_{n\in\mathcal{G}} q_n \sum_{x^n} V_n^N. \tag{2.51}$$

y en particular dicha cota se alcanza si se toma la regla de paro truncada con siguiente estructura:

Para $N \geq 1$ fijo y $\boldsymbol{n} = (n_1, n_2, \ldots, n_k) \in \mathcal{G}^k$,

$$\psi^N = \begin{cases} \psi_{\boldsymbol{n}} = I_{\left\{ l_{\boldsymbol{n}} \,\leq\, \overline{c} f_0^{\boldsymbol{n}} + \sum\limits_{m\in\mathcal{G}} q_m \sum\limits_{x_m^{(m)}} V_{(n,m)}^N \right\}}, & si\ 1 \leq k \leq N-1 \\ \\ \psi_{\boldsymbol{n}} = 1, & si\ k = N. \end{cases} \tag{2.52}$$

La regla de paro ψ^N truncada es la familia de funciones indicadoras cuyas componentes están definidas por $\psi_{\boldsymbol{n}}$, donde $\boldsymbol{n}$ es el vector de los tamaños consecutivos de los grupos observados en el experimento estadístico.

Teorema 3.

Sea ψ^N la regla de paro truncada cuyas componentes están definidas en (2.52), entonces

$$L_N\big(\psi^N\big) \;=\; \overline{c} + \sum_{\boldsymbol{n}\in\mathcal{G}} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} V_{\boldsymbol{n}}^N. \tag{2.53}$$

Demostración: Sea la regla de paro truncada ψ^N cuyas componentes están definidas en (2.52), entonces aplicando sucesivamente el Lema 3 para esta regla se obtiene la siguiente cadena de igualdades:

$$\begin{aligned}
L_N\big(\psi^N\big) &= \sum_{k=1}^{N-1}\sum_{\boldsymbol{n}\in\mathcal{G}^k}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi^N}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}}\Big) + \sum_{\boldsymbol{n}\in\mathcal{G}^N}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi^N}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}} + V_{\boldsymbol{n}}^N\Big)\\
&= \sum_{k=1}^{N-2}\sum_{\boldsymbol{n}\in\mathcal{G}^k}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi^N}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}}\Big) + \sum_{\boldsymbol{n}\in\mathcal{G}^{N-1}}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi^N}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}} + V_{\boldsymbol{n}}^N\Big)\\
&\;\;\vdots\\
&= \sum_{k=1}^{1}\sum_{\boldsymbol{n}\in\mathcal{G}^k}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi^N}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}}\Big) + \sum_{\boldsymbol{n}\in\mathcal{G}^2}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi^N}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}} + V_{\boldsymbol{n}}^N\Big)\\
&= \sum_{\boldsymbol{n}\in\mathcal{G}}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi^N}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}} + V_{\boldsymbol{n}}^N\Big)\\
&= \overline{c} + \sum_{\boldsymbol{n}\in\mathcal{G}} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} V_{\boldsymbol{n}}^N.
\end{aligned}$$

$\therefore$

$$L_N\big(\psi^N\big) = \overline{c} + \sum_{\boldsymbol{n}\in\mathcal{G}} q_n \sum_{x^{\boldsymbol{n}}} V_{\boldsymbol{n}}^N.$$

∎

Por lo tanto, ψ^N es la regla de paro truncada óptima porque alcanza la cota mínima de la función de Lagrange truncada.

2.2.3. Paro óptimo. Caso no truncado

En la sección anterior se fijó un tiempo de paro máximo N y se tomó la función de Lagrange $L_N(\psi)$ truncada con $\boldsymbol{n} \in \mathcal{G}^N$ (2.30), se acotó ésta mediante el Lema 2 y después se encontró la función que alcanza la cota mínima (2.53). En esta sección utilizaremos la función $L_N(\psi)$ y analizaremos el caso límite cuando N tiende a infinito, éste será *el caso no truncado*. Con ello encontraremos la *forma* de la regla de paro óptima general y resolveremos el problema de la prueba óptima (ψ^*, ϕ^*) que minimiza el costo.

Para una regla de paro ψ cualquiera definamos la regla de paro truncada ψ^{T_N} (T de truncada) asociada a ella. Para la cual se aplican sólo sus primeras $N-1$ componentes $\psi_{(\boldsymbol{n}:1)}, \psi_{(\boldsymbol{n}:2)}, \ldots, \psi_{(\boldsymbol{n}:N-1)}$ con $\boldsymbol{n} \in \mathcal{G}^N$ y para la componente $N-$ésima se tiene que $\psi_{\boldsymbol{n}} \equiv 1$. Debido a que toda regla de paro tiene asociada su regla truncada ψ^{T_N} es ahora a esta última a la que se le puede aplicar la función de Lagrange truncada dada en (2.30) y tener así $L_N(\psi^{T_N})$.
Entonces, para cualquier regla de paro ψ, $L_N(\psi)$ se define como:

$$L_N(\psi) = L_N\big(\psi^{T_N}\big) \tag{2.54}$$

Es fácil ver que $L_N(\psi)$ se puede calcular usando la fórmula (2.54) para cualquier regla de paro ψ no necesariamente truncada.

Lema 4.

Para cualquier regla de paro ψ tal que $\psi \in \Psi_{\theta_0}$,

$$\lim_{N\to\infty} L_N(\psi) = L(\psi). \tag{2.55}$$

Demostración.

Supongamos que $\psi \in \Psi_{\theta_0}$ y supongamos que $L(\psi) < \infty$. Después se analizará el caso cuando $L(\psi) = \infty$. Entonces calculemos la diferencia entre $L(\psi)$ y $L_N(\psi)$ y se demostrará que ésta tiende a cero cuando $N \to \infty$.

Como $\psi \in \Psi_{\theta_0}$, entonces

$$L(\psi) = \sum_{k=1}^{\infty} \sum_{\boldsymbol{n}\in\mathcal{G}^k} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s^{\psi}_{\boldsymbol{n}} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big)$$

$$\begin{aligned}
&L(\psi) - L_N(\psi) = \\
&= \sum_{k=1}^{\infty} \sum_{\boldsymbol{n}\in\mathcal{G}^k} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s^{\psi}_{\boldsymbol{n}} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big) - \\
&\quad - \left(\sum_{k=1}^{N-1} \sum_{\boldsymbol{n}\in\mathcal{G}^k} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s^{\psi}_{\boldsymbol{n}} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big) + \sum_{\boldsymbol{n}\in\mathcal{G}^N} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t^{\psi}_{\boldsymbol{n}} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big) \right) \\
&= \sum_{k=N}^{\infty} \sum_{\boldsymbol{n}\in\mathcal{G}^k} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s^{\psi}_{\boldsymbol{n}} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big) - \sum_{\boldsymbol{n}\in\mathcal{G}^N} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t^{\psi}_{\boldsymbol{n}} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big) \qquad (2.56)
\end{aligned}$$

Por hipótesis se tiene que $L(\psi) < \infty$ por tanto el primer término de (2.56) tiende a cero cuando $N \to \infty$ por ser la cola de una serie convergente. Analicemos el segundo termino.

$$\sum_{\boldsymbol{n}\in\mathcal{G}^N} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t^{\psi}_{\boldsymbol{n}} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big) = \sum_{\boldsymbol{n}\in\mathcal{G}^N} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t^{\psi}_{\boldsymbol{n}} c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + \sum_{\boldsymbol{n}\in\mathcal{G}^N} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t^{\psi}_{\boldsymbol{n}} l_{\boldsymbol{n}}. \tag{2.57}$$

Como $l_n = \text{mín}\{\lambda_0 f_0^{\boldsymbol{n}}, \lambda_1 f_1^{\boldsymbol{n}}\}$ entonces:

$$\begin{aligned}\sum_{\boldsymbol{n}\in\mathcal{G}^N}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi} l_{\boldsymbol{n}} &\leq \lambda_0 \sum_{\boldsymbol{n}\in\mathcal{G}^N}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi} f_0^{\boldsymbol{n}} \\ &= \lambda_0\, P_{\theta_0}\big(\tau^{\psi} \geq N\big) \to 0\end{aligned}$$

cuando $N \to \infty$ por (1.12).

Por otro lado:

$$\begin{aligned}\sum_{k=N}^{\infty}\sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}} c(\boldsymbol{n}) E_{\theta_0} s_{\boldsymbol{n}}^{\psi} &\geq \sum_{k=N}^{\infty}\sum_{\boldsymbol{n}\in\mathcal{G}^k} c(\boldsymbol{n}:N) q_{\boldsymbol{n}} E_{\theta_0} s_{\boldsymbol{n}}^{\psi} \\ &= \sum_{\boldsymbol{n}\in\mathcal{G}^N} c(\boldsymbol{n}) q_{\boldsymbol{n}} E_{\theta_0} s_{\boldsymbol{n}}^{\psi} + \sum_{k=N+1}^{\infty}\sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}} c(\boldsymbol{n}:N) E_{\theta_0} s_{\boldsymbol{n}}^{\psi} \\ &= \sum_{\boldsymbol{n}\in\mathcal{G}^N} c(\boldsymbol{n}) q_{\boldsymbol{n}} E_{\theta_0} s_{\boldsymbol{n}}^{\psi} + \sum_{\boldsymbol{n}\in\mathcal{G}^N} c(\boldsymbol{n}) \sum_{k=1}^{\infty}\sum_{m\in\mathcal{G}^k} q_{\boldsymbol{n}} q_{\boldsymbol{m}} E_{\theta_0} s_{\boldsymbol{n},\boldsymbol{m}}^{\psi} \\ &= \sum_{\boldsymbol{n}\in\mathcal{G}^N} c(\boldsymbol{n}) q_{\boldsymbol{n}} E_{\theta_0} t_{\boldsymbol{n}}^{\psi} = \sum_{\boldsymbol{n}\in\mathcal{G}^N} q_{\boldsymbol{n}} c(\boldsymbol{n}) \sum_{x^{\boldsymbol{n}}} t_{\boldsymbol{n}}^{\psi} f_0^{\boldsymbol{n}}.\end{aligned}$$

y ya sabemos que el lado izquierdo de la desigualdad converge a cero cuando $N \to \infty$ (por ser cola de una serie convergente).

Por tanto el segundo miembro de la desigualdad (2.57) también converge a cero cuando $N \to \infty$.

Ahora supongamos que $L(\psi) = \infty$, esto es:

$$\sum_{k=1}^{\infty}\sum_{\boldsymbol{n}\in\mathcal{G}^k}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}}\Big) = \infty$$

$\Rightarrow$

$$L_N(\psi) \geq \sum_{k=1}^{N-1}\sum_{\boldsymbol{n}\in\mathcal{G}^k}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}}\Big) \to \infty$$

cuando $N \to \infty$. De lo cual $L_N(\psi) \to \infty$ cuando $N \to \infty$.

$$\therefore \quad \lim_{N\to\infty} L_N(\psi) = L(\psi) \qquad \blacksquare$$

Ahora se verá el comportamiento de las funciones $V_{\boldsymbol{n}}^N$ (2.49) trabajadas en el lema 3, cuando $N \to \infty$.

Lema 5.

Sea $\boldsymbol{n} \in \mathcal{G}^r$ *y* $1 \leq r \leq N$, *entonces:*

$$V_{\boldsymbol{n}}^N \geq V_{\boldsymbol{n}}^{N+1}. \tag{2.58}$$

Demostración.

Se trabajará inducción sobre r con $r = N, N-1, \ldots, 1$.

Sea $r = N$ entonces $\boldsymbol{n} \in \mathcal{G}^N$ y por (2.49) se tiene que $V_{\boldsymbol{n}}^N = l_{\boldsymbol{n}}$. Entonces

$$V_{\boldsymbol{n}}^{N+1} = \text{mín}\left\{l_{\boldsymbol{n}}, \bar{c}f_0^{\boldsymbol{n}} + \sum_{\boldsymbol{m}\in\mathcal{G}} q_{\boldsymbol{m}} \sum_{x_{\boldsymbol{m}}^{(m)}} V_{(\boldsymbol{n},\boldsymbol{m})}^{N+1}\right\} \leq l_{\boldsymbol{n}} = V_{\boldsymbol{n}}^N$$

$$\text{Por tanto para } n \in \mathcal{G}^N \qquad V_n^N \geq V_n^{N+1}$$

Ahora supongamos que la desigualdad (2.58) se cumple para alguna r, con $1 < r \leq N$. Es decir suponga que se tiene:

$$V_{\boldsymbol{n}}^N \geq V_{\boldsymbol{n}}^{N+1} \qquad \boldsymbol{n} \in \mathcal{G}^r$$

Por demostrar que la propiedad se cumple para $r-1$. Sea $n \in \mathcal{G}^{r-1}$

$$\begin{aligned} V_{\boldsymbol{n}}^{N} &= \text{mín}\Big\{l_{\boldsymbol{n}}, \bar{c}f_0^{\boldsymbol{n}} + \sum_{\boldsymbol{m}\in\mathcal{G}} q_{\boldsymbol{m}} \sum_{x_m^{(m)}} V_{(\boldsymbol{n},m)}^{N}\Big\} \\ &\geq \text{mín}\Big\{l_{\boldsymbol{n}}, \bar{c}f_0^{\boldsymbol{n}} + \sum_{\boldsymbol{m}\in\mathcal{G}} q_{\boldsymbol{m}} \sum_{x_m^{(m)}} V_{(\boldsymbol{n},m)}^{N+1}\Big\} = V_{\boldsymbol{n}}^{N+1} \end{aligned}$$

$$\therefore \qquad V_{\boldsymbol{n}}^{N} \geq V_{\boldsymbol{n}}^{N+1} \qquad \boldsymbol{n} \in \mathcal{G}^{r-1}$$

■

El Lema 5 nos muestra que las funciones $V_{\boldsymbol{n}}^{N}$ con $\boldsymbol{n}$ fijo y $\boldsymbol{n} \in \mathcal{G}^r$ forman una sucesión decreciente y acotada, por tal su límite existe. Entonces dado $r \geq 1$ fijo y $\boldsymbol{n} \in \mathcal{G}^r$, sea

$$V_{\boldsymbol{n}} = \lim_{N\to\infty} V_{\boldsymbol{n}}^{N}. \tag{2.59}$$

Lema 6.

Sea ψ cualquier regla de paro. Entonces para cualquier $r \geq 1$ se cumplen las siguientes desigualdades:

$$\begin{aligned} L(\psi) &\geq \sum_{k=1}^{r} \sum_{\boldsymbol{n}\in\mathcal{G}^k} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}}\Big) + \sum_{\boldsymbol{n}\in\mathcal{G}^{r+1}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}} + V_{\boldsymbol{n}}\Big) \\ &\geq \sum_{k=1}^{r-1} \sum_{\boldsymbol{n}\in\mathcal{G}^k} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}}\Big) + \sum_{\boldsymbol{n}\in\mathcal{G}^{r}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}} + V_{\boldsymbol{n}}\Big) \end{aligned} \tag{2.60}$$

donde

$$V_{\boldsymbol{n}} = \text{mín}\Big\{l_{\boldsymbol{n}}, \bar{c}f_0^{\boldsymbol{n}} + \sum_{\boldsymbol{m}\in\mathcal{G}} q_{\boldsymbol{m}} \sum_{x_m^{(m)}} V_{(\boldsymbol{n},m)}\Big\} \tag{2.61}$$

Demostración

Sea ψ cualquier regla de paro.

Si $\psi \notin \psi_{\theta_0}$ entonces por (1.26) su función de Lagrange $L(\psi)$ será infinita. Por tanto para este caso la desigualdad (2.60) se cumple para cualquier $r \geq 1$. El caso interesante es si ψ termina el experimento. Supongamos que $\psi \in \Psi_{\theta_0}$, entonces por lema 3 se tiene que para cualquier $1 \leq r \leq N-1$ se cumple:

$$\begin{aligned} L_N(\psi) &\geq \sum_{k=1}^{r} \sum_{\boldsymbol{n} \in \mathcal{G}^k} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big) + \sum_{\boldsymbol{n} \in \mathcal{G}^{r+1}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + V_{\boldsymbol{n}}^N \Big) \\ &\geq \sum_{k=1}^{r-1} \sum_{\boldsymbol{n} \in \mathcal{G}^k} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big) + \sum_{\boldsymbol{n} \in \mathcal{G}^{r}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + V_{\boldsymbol{n}}^N \Big) \end{aligned}$$

donde $V_{\boldsymbol{n}}^N = l_{\boldsymbol{n}}$ con $\boldsymbol{n} \in \mathcal{G}^N$ y para $r < N$, $\boldsymbol{n} \in \mathcal{G}^r$

$$V_{\boldsymbol{n}}^N = \min \Big\{ l_{\boldsymbol{n}}, \bar{c} f_0^{\boldsymbol{n}} + \sum_{\boldsymbol{m} \in \mathcal{G}} q_{\boldsymbol{m}} \sum_{x_m^{(m)}} V_{(\boldsymbol{n},m)}^N \Big\}$$

Ahora, tomemos el límite cuando $N \to \infty$, por Lema 4 se obtiene la desigualdad (2.60). Y por Lema 5 se obtiene la ecuación para V_n dada en (2.61). Lo cual demuestra nuestro Lema. ∎

Con base en este lema 6 y en el lema 3 tomemos la siguiente la regla de paro:

Sea la **regla de paro** ψ^*, cuyas componentes definidas para $k \geq 1$, $\boldsymbol{n} \in \mathcal{G}^k$ como:

$$\psi_{\boldsymbol{n}}^* = I_{\Big\{ l_{\boldsymbol{n}} \,\leq\, \bar{c} f_0^{\boldsymbol{n}} + \sum_{m \in \mathcal{G}} q_m \sum_{x_m^{(m)}} V_{(\boldsymbol{n},m)} \Big\}}. \tag{2.62}$$

La regla de paro ψ^* es la familia de funciones indicadoras cuyas componentes están definidas por $\psi^*_{\boldsymbol{n}}$, donde $\boldsymbol{n}$ es el vector de los tamaños consecutivos de los grupos observados en el experimento estadístico.

Teorema 4.

La regla de paro ψ^* cuyas componentes están definidas en (2.62) *pertenece a la clase Ψ_{θ_0} y es óptima, en el sentido de que dada cualquier regla de paro ψ se cumple que:*

$$L(\psi) \quad \geq \quad L(\psi^*). \tag{2.63}$$

Además,

$$L(\psi^*) = \bar{c} + \sum_{\boldsymbol{n} \in \mathcal{G}} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} V_{\boldsymbol{n}}. \tag{2.64}$$

Demostración.

Sean ψ^* la regla de paro con componentes dadas en (2.62) y ψ una regla de paro cualquiera.

Observemos que para $r = 1$ se tiene, por lema 6, la siguiente cota inferior:

$$\sum_{\boldsymbol{n} \in \mathcal{G}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t^{\psi}_{\boldsymbol{n}} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + V_{\boldsymbol{n}} \Big) \quad = \quad \bar{c} + \sum_{\boldsymbol{n} \in \mathcal{G}} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} V_{\boldsymbol{n}}.$$

Ahora, pongamos a ψ^* en el lado izquierdo de (2.60), obteniendo:

$$L(\psi^*) \quad \geq \quad \bar{c} + \sum_{\boldsymbol{n} \in \mathcal{G}} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} V_{\boldsymbol{n}}. \tag{2.65}$$

Ahora, para cualquier $r \geq 1$ fijo, pongamos a ψ^* en el lado derecho de (2.60), y por lema 2 se tiene la siguiente cadena de igualdades.

$$\begin{aligned}
& \sum_{k=1}^{r} \sum_{\boldsymbol{n} \in \mathcal{G}^k} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi^*} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big) + \sum_{\boldsymbol{n} \in \mathcal{G}^{r+1}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi^*} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + V_{\boldsymbol{n}} \Big) \\
= \quad & \sum_{k=1}^{r-1} \sum_{\boldsymbol{n} \in \mathcal{G}^k} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi^*} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big) + \sum_{\boldsymbol{n} \in \mathcal{G}^{r}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi^*} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + V_{\boldsymbol{n}} \Big) \\
\vdots \quad & \\
= \quad & \sum_{\boldsymbol{n} \in \mathcal{G}^k} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi^*} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big) + \sum_{\boldsymbol{n} \in \mathcal{G}^{2}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi^*} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + V_{\boldsymbol{n}} \Big) \\
= \quad & \bar{c} + \sum_{\boldsymbol{n} \in \mathcal{G}} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} V_{\boldsymbol{n}}.
\end{aligned} \tag{2.66}$$

Si se toma el límite cuando $r \to \infty$ a dichas igualdades, éstas convergen a la constante. Esto es:

$$\lim_{r \to \infty} \left(\sum_{k=1}^{r} \sum_{\boldsymbol{n} \in \mathcal{G}^k} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi^*} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big) + \sum_{\boldsymbol{n} \in \mathcal{G}^{r+1}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi^*} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + V_{\boldsymbol{n}} \Big) \right) =$$

$$= \bar{c} + \sum_{\boldsymbol{n} \in \mathcal{G}} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} V_{\boldsymbol{n}}. \tag{2.67}$$

Por otro lado, de la cadena (2.66) se tiene:

$$\begin{aligned}
& \sum_{k=1}^{r} \sum_{\boldsymbol{n} \in \mathcal{G}^k} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi^*} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big) \leq \\
& \leq \sum_{k=1}^{r} \sum_{\boldsymbol{n} \in \mathcal{G}^k} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi^*} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}} \Big) + \sum_{\boldsymbol{n} \in \mathcal{G}^{r+1}} \sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi^*} \Big(c(\boldsymbol{n}) f_0^{\boldsymbol{n}} + V_{\boldsymbol{n}} \Big)
\end{aligned} \tag{2.68}$$

Tomando el límite en esta desigualdad (2.72) cuando $r \to \infty$

$$\begin{aligned}
&\lim_{r\to\infty}\left(\sum_{k=1}^{r}\sum_{\boldsymbol{n}\in\mathcal{G}^k}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi^*}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}}\Big)\right) \leq \\
&\leq \lim_{r\to\infty}\left(\sum_{k=1}^{r}\sum_{\boldsymbol{n}\in\mathcal{G}^k}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} s_{\boldsymbol{n}}^{\psi^*}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}}\Big) + \sum_{\boldsymbol{n}\in\mathcal{G}^{r+1}}\sum_{x^{\boldsymbol{n}}} q_{\boldsymbol{n}} t_{\boldsymbol{n}}^{\psi^*}\Big(c(\boldsymbol{n})f_0^{\boldsymbol{n}} + V_{\boldsymbol{n}}\Big)\right) \\
&= \bar{c} + \sum_{\boldsymbol{n}\in\mathcal{G}} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} V_{\boldsymbol{n}}. \qquad (2.69)
\end{aligned}$$

Si $\psi^* \in \Psi_{\theta_0}$, entonces por (1.23) se sabe que el lado izquierdo de la desigualdad anterior es $L(\psi^*)$ por lo tanto:

$$L(\psi^*) \quad \leq \quad \bar{c} + \sum_{\boldsymbol{n}\in\mathcal{G}} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} V_{\boldsymbol{n}}. \qquad (2.70)$$

De (2.65) y (2.70) se tiene que: si $\psi^* \in \Psi_{\theta_0}$, entonces

$$L(\psi^*) = \bar{c} + \sum_{\boldsymbol{n}\in\mathcal{G}} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} V_{\boldsymbol{n}}. \qquad (2.71)$$

Por lo tanto si $\psi^* \in \Psi_{\theta_0}$, del lema 6 y (2.71) se tiene:

$$L(\psi) \quad \geq \quad L(\psi^*) = \bar{c} + \sum_{\boldsymbol{n}\in\mathcal{G}} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} V_{\boldsymbol{n}}.$$

Ahora demostraremos que $\psi^* \in \Psi_{\theta_0}$.

De la cadena anterior (2.66) se tiene para $r \geq 1$ fijo que:

$$\sum_{k=1}^{r-1}\sum_{\boldsymbol{n}\in\mathcal{G}^k} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} s_{\boldsymbol{n}}^{\psi^*}\big(c(\boldsymbol{n})f_0^{\boldsymbol{n}} + l_{\boldsymbol{n}}\big) + \sum_{\boldsymbol{n}\in\mathcal{G}^r} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} t_{\boldsymbol{n}}^{\psi^*}\big(c(\boldsymbol{n})f_0^{\boldsymbol{n}} + V_{\boldsymbol{n}}\big) = \bar{c} + \sum_{\boldsymbol{n}\in\mathcal{G}} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} V_{\boldsymbol{n}},$$

entonces

$$\sum_{\boldsymbol{n}\in\mathcal{G}^r} q_{\boldsymbol{n}} c(\boldsymbol{n}) E_{\theta_0} t_{\boldsymbol{n}}^{\psi^*} = \sum_{\boldsymbol{n}\in\mathcal{G}^r} q_{\boldsymbol{n}} c(\boldsymbol{n}) \sum_{x^{\boldsymbol{n}}} t_{\boldsymbol{n}}^{\psi^*} f_0^{\boldsymbol{n}} \leq \bar{c} + \sum_{\boldsymbol{n}\in\mathcal{G}} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} V_{\boldsymbol{n}} = M < \infty \qquad (2.72)$$

porque, por suposición, $\bar{c} < \infty$; además $\sum_{\boldsymbol{n}\in\mathcal{G}} q_{\boldsymbol{n}} \sum_{x^{\boldsymbol{n}}} V_{\boldsymbol{n}}$ es finita por ser una suma finita.

Sea $K > 0$ un número cualquiera, tomemos la probabilidad de que el costo de varias etapas está por debajo de cota dada K.

$$\begin{aligned}
P\big(c_{\nu_1} + \cdots + c_{\nu_r} \leq K\big) &= P\left(\frac{c_{\nu_1} + \cdots + c_{\nu_r} - r\bar{c}}{\sqrt{r}} \leq \frac{K - r\bar{c}}{\sqrt{r}}\right) \\
&\leq P\left(\frac{c_{\nu_1} + \cdots + c_{\nu_r} - r\bar{c}}{\sqrt{r}} \leq 1 - \sqrt{r}\bar{c}\right)
\end{aligned}$$

si r es suficientemente grande y

$$\leq P\left(\frac{|c_{\nu_1} + \cdots + c_{\nu_r} - r\bar{c}|}{\sqrt{r}} \geq \sqrt{r}\bar{c} - 1\right).$$

Por la desigualdad de Chebyshev

$$\leq \frac{\operatorname{Var} c_\nu}{(\sqrt{r}\bar{c} - 1)^2}.$$

Ahora tómese el límite cuando $r \to \infty$, entonces

$$\sum_{\boldsymbol{n} \in \mathcal{G}^r,\, c(\boldsymbol{n}) \leq K} q_{\boldsymbol{n}} = P\Big(c_{\nu_1} + \cdots + c_{\nu_r} \leq K\Big) \to 0, \tag{2.73}$$

para cualquier K positivo.

De (2.72) se sigue

$$K \sum_{\boldsymbol{n} \in \mathcal{G}^r,\, c(\boldsymbol{n}) > K} q_{\boldsymbol{n}} E_{\theta_0} t_{\boldsymbol{n}}^{\psi^*} < \sum_{\boldsymbol{n} \in \mathcal{G}^r,\, c(\boldsymbol{n}) > K} q_{\boldsymbol{n}} c(\boldsymbol{n}) E_{\theta_0} t_{\boldsymbol{n}}^{\psi^*} \leq M$$

para r suficientemente grande. Entonces

$$\sum_{\boldsymbol{n} \in \mathcal{G}^r} q_{\boldsymbol{n}} E_{\theta_0} t_{\boldsymbol{n}}^{\psi^*} - \sum_{\boldsymbol{n} \in \mathcal{G}^r,\, c(\boldsymbol{n}) \leq K} q_{\boldsymbol{n}} E_{\theta_0} t_{\boldsymbol{n}}^{\psi^*} \leq \frac{M}{K}$$

ó

$$\sum_{\boldsymbol{n} \in \mathcal{G}^r} q_{\boldsymbol{n}} E_{\theta_0} t_{\boldsymbol{n}}^{\psi^*} \leq \frac{M}{K} + \sum_{\boldsymbol{n} \in \mathcal{G}^r,\, c(\boldsymbol{n}) \leq K} q_{\boldsymbol{n}} E_{\theta_0} t_{\boldsymbol{n}}^{\psi^*} \leq \frac{M}{K} + \sum_{\boldsymbol{n} \in \mathcal{G}^r,\, c(\boldsymbol{n}) \leq K} q_{\boldsymbol{n}} \tag{2.74}$$

Sea $\varepsilon > 0$ cualquier número, y sea $K = \frac{M}{\varepsilon}$. De (2.74) tenemos

$$\sum_{\boldsymbol{n}\in\mathcal{G}^r} q_{\boldsymbol{n}} E_{\theta_0} t_{\boldsymbol{n}}^{\psi^*} \leq \varepsilon + \sum_{\boldsymbol{n}\in\mathcal{G}^r,\ c(\boldsymbol{n})\leq K} q_{\boldsymbol{n}},$$

y de (2.73) concluimos que para r suficientemente grande

$$\sum_{\boldsymbol{n}\in\mathcal{G}^r} q_{\boldsymbol{n}} E_{\theta_0} t_{\boldsymbol{n}}^{\psi^*} < 2\varepsilon.$$

Esto implica que

$$\lim_{r\to\infty} P_{\theta_0}(\tau^{\psi^*} \geq r) = \lim_{r\to\infty} \sum_{\boldsymbol{n}\in\mathcal{G}^N} q_{\boldsymbol{n}} E_{\theta_0} t_{\boldsymbol{n}}^{\psi^*} = 0,$$

$$\Rightarrow \quad P_{\theta_0}(\tau^{\psi^*} < \infty) = 1 \qquad \Rightarrow \quad \psi^* \in \Psi_{\theta_0}.$$

■

Hemos encontrado la prueba secuencial (ψ^*, ϕ^*) cuyas componentes están definidas mediante:

$$\begin{aligned} \psi^*_{\boldsymbol{n}} &= I_{\left\{l_{\boldsymbol{n}} \,\leq\, \bar{c} f_0^{\boldsymbol{n}} + \sum\limits_{\boldsymbol{m}\in\mathcal{G}} q_{\boldsymbol{m}} \sum\limits_{x_m^{(m)}} V_{(\boldsymbol{n},m)}\right\}} \\ \phi^*_{\boldsymbol{n}} &= I_{\left\{\lambda_0 f_0^{\boldsymbol{n}} \leq \lambda_1 f_1^{\boldsymbol{n}}\right\}} \end{aligned} \tag{2.75}$$

para $k \geq 1,\ \boldsymbol{n} \in \mathcal{G}^k$, con la propiedad de que $\psi^* \in \Psi_{\theta_0}$.

Finalmente el siguiente teorema demuestra la optimalidad de la prueba secuencial encontrada.

Teorema 5.

Dada la prueba secuencial (ψ^*, ϕ^*)*, cuyas respectivas componentes están dadas en* (2.75)*, entonces para cualquier prueba* (ψ, ϕ) *tal que:*

$$\alpha(\psi, \phi) \leq \alpha(\psi^*, \phi^*), \qquad \beta(\psi, \phi) \leq \beta(\psi^*, \phi^*)$$

se tiene que

$$K(\psi) \geq K(\psi^*) \tag{2.76}$$

Demostración

Sea (ψ, ϕ) una prueba secuencial cualquiera y sea (ψ^*, ϕ^*) la prueba cuyas componentes están definidas en (2.75). Por teorema 4 se tiene que:

$$L(\psi) \geq L(\psi^*) \quad \Rightarrow \quad L(\psi, \phi^*) \geq L(\psi^*, \phi^*)$$

Pero por teorema 2 (2.15), para cualquier ϕ

$$L(\psi, \phi) \geq L(\psi^*, \phi^*),$$

donde ψ es arbitraria.

Ahora sea $\alpha(\psi^*, \phi^*) = \alpha$ y $\beta(\psi^*, \phi^*) = \beta$, entonces aplicando el teorema 1 se tiene que $K(\psi) \geq K(\psi^*)$. ■

Con base en el análisis anterior hemos encontrado la prueba secuencial óptima (ψ^*, ϕ^*) que conlleva a un costo promedio mínimo (bajo la hipótesis nula), entre todas las pruebas con probabilidades de error menores o iguales que las de (ψ^*, ϕ^*).

2.2.4. Estructura de las pruebas secuenciales óptimas

En las secciones previas se ha demostrado que tanto la prueba secuencial (ψ^N, ϕ^*) que contiene a la regla paro truncada, como la prueba secuencial (ψ^*, ϕ^*) son óptimas en su clase respectiva; ya que conllevan a una función de Lagrange menor y por tal a un costo mínimo.

Para recapitular se presentan las estructuras de estas pruebas secuenciales óptimas.

La prueba secuencial con regla de paro truncada es (ψ^N, ϕ^*) cuyas componentes son:

$$\psi^N = \begin{cases} \psi_{\boldsymbol{n}} = I_{\left\{ l_{\boldsymbol{n}} \leq \bar{c} f_0^{\boldsymbol{n}} + \sum\limits_{\boldsymbol{m} \in \mathcal{G}} q_{\boldsymbol{m}} \sum\limits_{x_{\boldsymbol{m}}^{(m)}} V_{(\boldsymbol{n},\boldsymbol{m})}^N \right\}}, & si\ 1 \leq k \leq N-1 \\ \psi_{\boldsymbol{n}} = 1, & si\ k = N. \end{cases}$$

$$\phi_{\boldsymbol{n}}^* = I_{\left\{ \lambda_0 f_0^{\boldsymbol{n}} \leq \lambda_1 f_1^{\boldsymbol{n}} \right\}}.$$

La estructura de la prueba secuencial óptima (ψ^*, ϕ^*) es:

$$\psi_{\boldsymbol{n}}^* = I_{\left\{ l_{\boldsymbol{n}} \leq \bar{c} f_0^{\boldsymbol{n}} + \sum\limits_{\boldsymbol{m} \in \mathcal{G}} q_{\boldsymbol{m}} \sum\limits_{x_{\boldsymbol{m}}^{(m)}} V_{(\boldsymbol{n},\boldsymbol{m})} \right\}}.$$

$$\phi_{\boldsymbol{n}}^* = I_{\left\{ \lambda_0 f_0^{\boldsymbol{n}} \leq \lambda_1 f_1^{\boldsymbol{n}} \right\}}.$$

Observación: Tómese en cuenta que hasta este momento en nuestro trabajo NO hemos utilizado la independencia de las observaciones, aún cuando estamos suponiendo este hecho. Entonces, podemos decir que todo el trabajo anterior (casos de paro truncado y no truncado) es aplicable a situaciones más generales, en las cuales la independencia de las observaciones no se tenga.

CAPÍTULO 3

LA PRUEBA SECUENCIAL ALEATORIA DE LA RAZÓN DE PROBABILIDADES (RSPRT)

3.1. La prueba óptima en términos de la razón de probabilidades

En esta sección se mostrará que la prueba secuencial óptima, con componentes dadas en (2.75), obtenida en sección anterior es equivalente a la conocida *prueba secuencial de razón de probabilidades aleatoria*, (*RSPRT*, por sus siglas en inglés: *Random Sequential Probability Ratio Test*). Para lo cual se expresarán las componentes de la prueba óptima en términos de la razón de probabilidades.

Supongamos que $f_{\theta_0}(x) > 0$, $f_{\theta_1}(x) > 0$ para toda $x \in \mathcal{X}$ y además que existe al menos una $x \in \mathcal{X}$ para la cual $f_{\theta_0}(x) \neq f_{\theta_1}(x)$, *i. e.* supóngase que las respectivas distribuciones bajo las hipótesis H_0 y H_1 son diferentes, porque no tendría sentido práctico realizar una prueba de hipótesis cuando las distribuciones a contrastar son iguales. También sea $\boldsymbol{n} = (n_1, \ldots, n_k)$, k grupos con sus respectivas observaciones $x^{\boldsymbol{n}} = \big(x_1^{(n_1)}, \ldots, x_k^{(n_k)}\big)$, ahora tomemos la razón de probabilidades (conjuntas) mediante:

$$Z_{\boldsymbol{n}} = \frac{\prod_{i=1}^{k}\prod_{j=1}^{n_i} f_{\theta_1}(x_{ij})}{\prod_{i=1}^{k}\prod_{j=1}^{n_i} f_{\theta_0}(x_{ij})} = \frac{f_1^{\boldsymbol{n}}}{f_0^{\boldsymbol{n}}} \quad \text{con } \boldsymbol{n} \in \mathcal{G}^k,\ k = 1, 2, \ldots. \tag{3.1}$$

Como existe al menos un $x \in \mathcal{X}$ tal que $f_{\theta_0}(x) \neq f_{\theta_1}(x)$ entonces:

$$P_{\theta_0}\big(Z_{\boldsymbol{n}} = 1\big) < 1 \tag{3.2}$$

Definamos una función g que depende de z mediante:

$$g(z) \quad = \quad \text{mín}\{\lambda_0, \lambda_1 z\},\ z \in \mathbb{R}^+. \tag{3.3}$$

Vamos a trabajar con las funciones $V_{\boldsymbol{n}}^N$ (2.49), para lo cual se formarán unas nuevas funciones llamadas $U_{\boldsymbol{n}}^N$, inductivamente por medio de la dimensión r de $\boldsymbol{n}$.

Primero tómese $r = N$ entonces $\boldsymbol{n} \in \mathcal{G}^N$, se sabe que:

$$\begin{aligned} V_{\boldsymbol{n}}^N &= l_{\boldsymbol{n}} \\ &= \text{mín}\{\lambda_0 f_0^{\boldsymbol{n}}, \lambda_1 f_1^{\boldsymbol{n}}\} \\ &= \text{mín}\left\{\lambda_0, \lambda_1 \frac{f_1^{\boldsymbol{n}}}{f_0^{\boldsymbol{n}}}\right\} f_0^{\boldsymbol{n}} \\ V_{\boldsymbol{n}}^N &= g(Z_{\boldsymbol{n}})\ f_0^{\boldsymbol{n}} \end{aligned} \tag{3.4}$$

Ahora para $\boldsymbol{n} \in \mathcal{G}^N$ sea

$$U_{\boldsymbol{n}}^N(Z_{\boldsymbol{n}}) \;=\; g(Z_{\boldsymbol{n}}). \tag{3.5}$$

Sustituyendo y despejando en la ecuación (3.4) se obtiene que:

$$U_{\boldsymbol{n}}^N(Z_{\boldsymbol{n}}) \;=\; \frac{V_{\boldsymbol{n}}^N}{f_0^{\boldsymbol{n}}} \qquad \text{con } \boldsymbol{n} \in \mathcal{G}^N \tag{3.6}$$

Para $r = N-1$ y $\boldsymbol{n} \in \mathcal{G}^{N-1}$, véase (2.40):

$$\begin{aligned}
V_{\boldsymbol{n}}^N &= \min\left\{ l_{\boldsymbol{n}},\ \bar{c} f_0^{\boldsymbol{n}} + \sum_{\boldsymbol{m}\in\mathcal{G}} q_{\boldsymbol{m}} \sum_{x_m^{(m)}} V_{(\boldsymbol{n},m)}^N \right\} \\
&= \min\left\{ \min\{\lambda_0 f_0^{\boldsymbol{n}}, \lambda_1 f_1^{\boldsymbol{n}}\},\ \bar{c} f_0^{\boldsymbol{n}} + \sum_{m\in\mathcal{G}} q_m \sum_{x_m^{(m)}} V_{(\boldsymbol{n},m)}^N \right\} \\
&= \min\left\{ \min\left\{\lambda_0, \lambda_1 \frac{f_1^{\boldsymbol{n}}}{f_0^{\boldsymbol{n}}}\right\},\ \bar{c} + \sum_{m\in\mathcal{G}} q_m \sum_{x_m^{(m)}} \frac{V_{(\boldsymbol{n},m)}^N}{f_0^{\boldsymbol{n}}} \right\} f_0^{\boldsymbol{n}} \\
&= \min\left\{ g(Z_{\boldsymbol{n}}),\ \bar{c} + \sum_{m\in\mathcal{G}} q_m \sum_{x_m^{(m)}} \frac{V_{(\boldsymbol{n},m)}^N}{f_0^{(\boldsymbol{n},m)}} f_0^m\big(x_m^{(m)}\big) \right\} f_0^{\boldsymbol{n}} \\
&= \min\left\{ g(Z_{\boldsymbol{n}}),\ \bar{c} + \sum_{m\in\mathcal{G}} q_m \sum_{x_m^{(m)}} U_{(\boldsymbol{n},m)}^N\big(Z_{(\boldsymbol{n},m)}\big) f_0^m\big(x_m^{(m)}\big) \right\} f_0^{\boldsymbol{n}}.
\end{aligned}$$

Ahora tomemos en cuenta que para $\boldsymbol{n} \in \mathcal{G}^{N-1}$, el vector $(\boldsymbol{n}, m) \in \mathcal{G}^N$ entonces:

$$\begin{aligned}
Z_{(\boldsymbol{n},m)} &= \frac{\prod_{i=1}^{N-1}\prod_{j=1}^{n_i} f_{\theta_1}(x_{ij})}{\prod_{i=1}^{N-1}\prod_{j=1}^{n_i} f_{\theta_0}(x_{ij})}\ \frac{\prod_{j=1}^{m} f_{\theta_1}(x_{mj})}{\prod_{j=1}^{m} f_{\theta_0}(x_{mj})} \\
&= Z_{\boldsymbol{n}} \prod_{j=1}^{m} \frac{f_{\theta_1}(x_{mj})}{f_{\theta_0}(x_{mj})} = Z_{\boldsymbol{n}} \frac{f_{\theta_1}^m\big(x_m^{(m)}\big)}{f_{\theta_0}^m\big(x_m^{(m)}\big)}.
\end{aligned}$$

Donde $f_{\theta_i}^m\big(x_m^{(m)}\big)$ para $i=0,1$ es, respectivamente, la función de probabilidad conjunta aplicada a los elementos del último grupo de tamaño m. Ahora sustituyendo esta última igualdad en $V_{\boldsymbol{n}}^N$ con $\boldsymbol{n}\in\mathcal{G}^{N-1}$ anterior se tiene:

$$V_{\boldsymbol{n}}^N \;=\; \text{mín}\left\{g(Z_{\boldsymbol{n}}),\ \bar{c}+\sum_{m\in\mathcal{G}} q_m \sum_{x_m^{(m)}} f_0^m\big(x_m^{(m)}\big)U_{(\boldsymbol{n},m)}^N\left(Z_{\boldsymbol{n}}\frac{f_{\theta_1}^m\big(x_m^{(m)}\big)}{f_{\theta_0}^m\big(x_m^{(m)}\big)}\right)\right\}f_0^{\boldsymbol{n}}$$

Ahora para $\boldsymbol{n}\in\mathcal{G}^{N-1}$ sea

$$U_{\boldsymbol{n}}^N(Z_{\boldsymbol{n}}) \;=\; \text{mín}\left\{g(Z_{\boldsymbol{n}}),\ \bar{c}+\sum_{m\in\mathcal{G}} q_m \sum_{x_m^{(m)}} f_0^m\big(x_m^{(m)}\big)U_{(\boldsymbol{n},m)}^N\left(Z_{\boldsymbol{n}}\frac{f_{\theta_1}^m\big(x_m^{(m)}\big)}{f_{\theta_0}^m\big(x_m^{(m)}\big)}\right)\right\}. \tag{3.7}$$

Por tanto

$$V_{\boldsymbol{n}}^N = U_{\boldsymbol{n}}^N(Z_n)f_0^{\boldsymbol{n}} \qquad \Rightarrow \qquad U_{\boldsymbol{n}}^N(Z_{\boldsymbol{n}}) = \frac{V_{\boldsymbol{n}}^N}{f_0^{\boldsymbol{n}}} \quad \text{con } \boldsymbol{n}\in\mathcal{G}^{N-1}$$

Con base en los pasos anteriores podemos relacionar las funciones $U_{\boldsymbol{n}}^N(Z_{\boldsymbol{n}}),\ \ \boldsymbol{n}\in\mathcal{G}^r$ en general para $1\le r\le N$ como:

$$U_{\boldsymbol{n}}^N(Z_{\boldsymbol{n}}) = \frac{V_{\boldsymbol{n}}^N}{f_0^{\boldsymbol{n}}} \qquad \text{con } \boldsymbol{n}\in\mathcal{G}^r \tag{3.8}$$

Donde éstas por definición son:

$$U_{\boldsymbol{n}}^N(Z_{\boldsymbol{n}}) \;=\; \text{mín}\left\{g(Z_{\boldsymbol{n}}),\ \bar{c}+\sum_{m\in\mathcal{G}} q_m \sum_{x_m^{(m)}} f_0^m\big(x_m^{(m)}\big)U_{(\boldsymbol{n},m)}^N\left(Z_{\boldsymbol{n}}\frac{f_{\theta_1}^m\big(x_m^{(m)}\big)}{f_{\theta_0}^m\big(x_m^{(m)}\big)}\right)\right\} \tag{3.9}$$

Se observa explícitamente, en esta última relación que las funciones dependen de la razón de probabilidades $Z_{\boldsymbol{n}}$ (3.1).

Ahora vamos a formar una sucesión de funciones llamadas ρ_k de la siguiente manera:

$$\rho_0(z) = g(z) \tag{3.10}$$

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

$$\begin{aligned}\rho_k(z) &= \text{mín}\left\{g(z), \bar{c} + \sum_{m\in\mathcal{G}} q_m \sum_{x_m^{(m)}} \rho_{k-1}\left(z\frac{f_1^m\left(x_m^{(m)}\right)}{f_0^m\left(x_m^{(m)}\right)}\right)\cdot f_0^m\left(x_m^{(m)}\right)\right\}\\ \Rightarrow \quad & \\ \rho_k(z) &= \text{mín}\left\{g(z), \bar{c} + \sum_{m\in\mathcal{G}} q_m E_{\theta_0}\rho_{k-1}\left(z\frac{f_1^m\left(X_m^{(m)}\right)}{f_0^m\left(X_m^{(m)}\right)}\right)\right\}. \end{aligned} \tag{3.11}$$

Se puede demostrar fácilmente que la sucesión $\{\rho_k(z)\}_k$ es decreciente para una z fija y no negativa ($z \geq 0$); entonces esta sucesión es acotada; por tanto es convergente para cada $z \in \mathbb{R}^+$. Sea por definición,

$$\rho(z) = \lim_{k\to\infty} \rho_k(z). \tag{3.12}$$

Ahora tomando el límite a la ecuación (3.11), cuando $k \to \infty$ se obtiene:

$$\rho(z) = \text{mín}\left\{g(z), \bar{c} + \sum_{m\in\mathcal{G}} q_m E_{\theta_0}\rho\left(z\frac{f_1^m\left(X_m^{(m)}\right)}{f_0^m\left(X_m^{(m)}\right)}\right)\right\} \tag{3.13}$$

Esta ecuación es llamada la *ecuación de la programación dinámica* (o ecuación de Bellman).

Con (3.10) y (3.11) se define una sucesión cuyos elementos dependen únicamente de z. En particular para el caso truncado ($n \in \mathcal{G}^N$) y utilizando las definiciones (3.5), (3.7) y (3.9), se puede ver fácilmente que se obtienen N de elementos de la sucesión ρ_k, los cuales dependen de la razón de probabilidades Z_n, a saber:

$$\begin{aligned}\rho_0(Z_{\boldsymbol{n}}) &= U_{\boldsymbol{n}}^N(Z_{\boldsymbol{n}}) \quad \text{con } \boldsymbol{n} \in \mathcal{G}^N\\ \rho_1(Z_{\boldsymbol{n}}) &= U_{\boldsymbol{n}}^N(Z_{\boldsymbol{n}}) \quad \text{con } \boldsymbol{n} \in \mathcal{G}^{N-1}\\ &\vdots\\ \rho_k(Z_{\boldsymbol{n}}) &= U_{\boldsymbol{n}}^N(Z_{\boldsymbol{n}}) \quad \text{con } \boldsymbol{n} \in \mathcal{G}^{N-k}\\ &\vdots\\ \rho_{N-1}(Z_{\boldsymbol{n}}) &= U_{\boldsymbol{n}}^N(Z_{\boldsymbol{n}}) \quad \text{con } \boldsymbol{n} \in \mathcal{G}\end{aligned}$$

Naturalmente:

$$U_{\boldsymbol{n}}(Z_{\boldsymbol{n}}) = \lim_{N\to\infty} U_{\boldsymbol{n}}^N(Z_{\boldsymbol{n}}) = \lim_{N\to\infty} \rho_{N-k}(Z_{\boldsymbol{n}}) = \rho(Z_{\boldsymbol{n}}) \quad \text{con } \boldsymbol{n} \in \mathcal{G}^k. \tag{3.14}$$

Por otro lado, de la sección anterior se sabe, por (2.59), que:

$$V_{\boldsymbol{n}} = \lim_{N\to\infty} V_{\boldsymbol{n}}^N.$$

entonces

$$U_{\boldsymbol{n}} = \lim_{N\to\infty} U_{\boldsymbol{n}}^N = \lim_{N\to\infty} \frac{V_{\boldsymbol{n}}^N}{f_0^{\boldsymbol{n}}} = \frac{V_{\boldsymbol{n}}}{f_0^{\boldsymbol{n}}}. \tag{3.15}$$

Ahora escribamos la regla de paro óptima ψ^* de las ecuaciones (2.75) en términos de lo anterior.

$$\begin{aligned} \psi_{\boldsymbol{n}}^* &= I_{\left\{ l_{\boldsymbol{n}} \leq \bar{c} f_0^{\boldsymbol{n}} + \sum_{m\in\mathcal{G}} q_m \sum_{x_m^{(m)}} V_{(\boldsymbol{n},m)} \right\}} \\ &= I_{\left\{ g(Z_{\boldsymbol{n}}) \leq \bar{c} + \sum_{m\in\mathcal{G}} q_m E_{\theta_0} \rho\left(Z_{\boldsymbol{n}} \frac{f_1^m\left(X_m^{(m)}\right)}{f_0^m\left(X_m^{(m)}\right)} \right) \right\}} \end{aligned} \tag{3.16}$$

Analicemos ahora las características de las funciones que ocurren en la desigualdad:

$$g(z) \leq \bar{c} + \sum_{m\in\mathcal{G}} q_m E_{\theta_0} \rho\left(z \frac{f_1^m\left(X_m^{(m)}\right)}{f_0^m\left(X_m^{(m)}\right)} \right) \tag{3.17}$$

Lema 7.

Las funciones $\rho_k(z)$ definidas en (3.10)-(3.11) *y la función $\rho(z)$* (3.13) *todas poseen las siguientes propiedades:*

a) $\rho_k(0) = \rho(0) = 0$;

b) *son cóncavas y continuas para todo $z \in [0, \infty)$;*

c) *son crecientes en el intervalo $[0, \infty)$; y*

d) $\lim\limits_{z\to\infty} \rho_k(z) = \lim\limits_{z\to\infty} \rho(z) = \lambda_0$.

Demostración

a) El hecho de que $\rho_k(0) = 0 \quad \forall k \geq 0$ es fácilmente demostrable por inducción. Por tanto $\rho(0) = \lim\limits_{k\to\infty} \rho_k(0) = 0$.

b) Demostremos también por inducción que $\rho_k(z)$ son cóncavas.

Para $k = 0$

$\rho_0(z) = g(z)$ es cóncava por ser el mínimo de dos funciones cóncavas (en este caso en particular funciones lineales).

Sea $n \in \mathbb{N}$ fijo supongamos que $\rho_k(z)$ es cóncava para toda $k \leq n$,

Ahora *demostraremos* que las funciones $E_{\theta_0}\rho_k\left(z\frac{f_1^m\left(X_m^{(m)}\right)}{f_0^m\left(X_m^{(m)}\right)}\right)$ son cóncavas $\forall\ k \leq n$.

En efecto, por hipótesis sabemos que $\rho_k(z)$ es cóncava para toda $k \leq n$, esto es por definición:

$\forall z_1, z_2 \in [0, \infty)$ y $\forall \alpha \in [0,1]$,

$$\alpha\rho_k(z_1) + (1-\alpha)\rho_k(z_2) \quad \leq \quad \rho_k\Big(\alpha z_1 + (1-\alpha)z_2\Big).$$

Sea un valor $x_m^{(m)}$ cualquiera, entonces:

$$\begin{aligned}
&\alpha\rho_k\left(z_1\frac{f_1^m\big(x_m^{(m)}\big)}{f_0^m\big(x_m^{(m)}\big)}\right) + (1-\alpha)\rho_k\left(z_2\frac{f_1^m\big(x_m^{(m)}\big)}{f_0^m\big(x_m^{(m)}\big)}\right)\\
&\leq \rho_k\left(\Big(\alpha z_1 + (1-\alpha)z_2\Big)\frac{f_1^m\big(x_m^{(m)}\big)}{f_0^m\big(x_m^{(m)}\big)}\right).
\end{aligned}$$

Esta desigualdad se cumple para cualquiera $x_m^{(m)}$, entonces:

$$\begin{aligned}
&\sum_{x_m^{(m)}}\alpha\rho_k\left(z_1\frac{f_1^m\big(x_m^{(m)}\big)}{f_0^m\big(x_m^{(m)}\big)}\right) f_0^m\big(x_m^{(m)}\big) + \sum_{x_m^{(m)}}(1-\alpha)\rho_k\left(z_2\frac{f_1^m\big(x_m^{(m)}\big)}{f_0^m\big(x_m^{(m)}\big)}\right) f_0^m\big(x_m^{(m)}\big)\\
&\leq \sum_{x_m^{(m)}}\rho_k\left(\Big(\alpha z_1 + (1-\alpha)z_2\Big)\frac{f_1^m\big(x_m^{(m)}\big)}{f_0^m\big(x_m^{(m)}\big)}\right)f_0^m\big(x_m^{(m)}\big).
\end{aligned}$$

Por definición y propiedades del valor esperado, se obtiene:

$$\begin{aligned}
&\alpha E_{\theta_0}\rho_k\left(z_1\frac{f_1^m\big(X_m^{(m)}\big)}{f_0^m\big(X_m^{(m)}\big)}\right) + (1-\alpha)E_{\theta_0}\rho_k\left(z_2\frac{f_1^m\big(X_m^{(m)}\big)}{f_0^m\big(X_m^{(m)}\big)}\right)\\
&\leq E_{\theta_0}\rho_k\left(\Big(\alpha z_1 + (1-\alpha)z_2\Big)\frac{f_1^m\big(X_m^{(m)}\big)}{f_0^m\big(X_m^{(m)}\big)}\right).
\end{aligned}$$

$\therefore$ las funciones $E_{\theta_0}\rho_k\left(z\frac{f_1^m\left(X_m^{(m)}\right)}{f_0^m\left(X_m^{(m)}\right)}\right)$ son cóncavas $\forall\ k \leq n$. Entonces por (3.11) ρ_{n+1} es el mínimo entre una función cóncava $g(z)$ y una suma de funciones cóncavas [14]. Por tanto ρ_{n+1} es también cóncava.

De lo cual $\rho_k(z)$ es cóncava para toda $k \in \mathbb{N}$

El límite puntual de una sucesión de funciones cóncavas es una función cóncava [14], así que $\rho(z) = \lim\limits_{k\to\infty} \rho_k(z)$ es también cóncava.
Como $\rho_k(z) \geq 0$, $\rho(z) \geq 0$ (éste por ser el límite de funciones no negativas) y además todas son cóncavas, entonces $\forall\ k\ \rho_k(z),\ \rho(z)$ son continuas en $(0, \infty)$.

De ***(a)*** se sabe que $\rho_k(0) = 0$ para toda $k \geq 0$, ahora para k fija se sabe que:

$$\begin{aligned} \rho_k(z) &\leq g(z) \\ \Rightarrow \quad \lim_{z\to 0^+} \sup_k \rho_k(z) &\leq \lim_{z\to 0^+} g(z) = 0 \end{aligned}$$

Por otro lado $\rho_k(z) \geq 0$, así que $\lim\limits_{z\to 0^+} \rho_k(z) = 0 = \rho_k(0)$.
Por tanto $\rho_k(z)$ es continua en 0, para toda $k \geq 0$.
Análogamente $\rho(z) \leq g(z) \to 0 = \rho(0)$, cuando $z \to 0$.
Por lo tanto las funciones $\rho_k(z),\ \rho(z)$ son continuas en $[0, \infty)$.

c) También demostremos por inducción que ρ_k son crecientes.
Para $k = 0$,

$$\rho_0(z) = \min\{\lambda_0, \lambda_1 z\} \quad \text{es creciente.}$$

Sea n un número natural fijo supongamos que $\rho_k(z)$ es creciente para toda $k \leq n$.
De lo cual y por (3.11) $\rho_{n+1}(z)$ es el mínimo entre dos funciones crecientes. Por tanto $\rho_{n+1}(z)$ es creciente.
El límite puntual de funciones crecientes es también creciente, por tal $\rho(z)$ es creciente.
Por lo tanto las funciones $\rho_k(z),\ \rho(z)$ son crecientes en el intervalo $[0, \infty)$.

d) Por las propiedades anteriores y utilizando, nuevamente, inducción demostraremos

que: $\lim\limits_{z\to\infty} \rho_k(z) = \lambda_0 \ \forall k \geq 0$.
Nótese primero que: $\lim\limits_{z\to\infty} g(z) = \lim\limits_{z\to\infty} \big(\text{mín}\{\lambda_0, \lambda_1 z\}\big) = \lambda_0$.
Para $k = 0$,

$$\lim_{z\to\infty} \rho_0(z) = \lim_{z\to\infty} g(z) = \lambda_0.$$

Ahora sea $n \in \mathbb{N}$ fijo, supongamos que $\lim\limits_{z\to\infty} \rho_k(z) = \lambda_0$ para toda $k \leq n$. Entonces $\rho_{n+1}(z)$ se define:

$$\rho_{n+1}(z) \quad = \quad \text{mín}\left\{g(z), \overline{c} + \sum_{m\in\mathcal{G}} q_m E_{\theta_0}\rho_n\left(z\frac{f_1^m\big(X_m^{(m)}\big)}{f_0^m\big(X_m^{(m)}\big)}\right)\right\}$$

Tomando el límite a ésta ecuación cuando $z \to \infty$:

$$\begin{aligned}
\lim_{z\to\infty} \rho_{n+1}(z) &= \lim_{z\to\infty}\left[\text{mín}\left\{g(z), \overline{c} + \sum_{m\in\mathcal{G}} q_m E_{\theta_0}\rho_n\left(z\frac{f_1^m\big(X_m^{(m)}\big)}{f_0^m\big(X_m^{(m)}\big)}\right)\right\}\right] \\
&= \text{mín}\{\lambda_0, \overline{c} + \lambda_0\} \\
&= \lambda_0
\end{aligned}$$

Por tanto $\forall \ k \geq 0$, $\lim\limits_{z\to\infty} \rho_k(z) = \lambda_0$.

Por **c)** el límite $\lim\limits_{z\to\infty} \rho(z) = \lambda$ existe. Ahora pasando al límite en la ecuación (3.13), cuando $z \to \infty$ se tiene:

$$\begin{aligned}
\lim_{z\to\infty} \rho(z) &= \lim_{z\to\infty}\left(\text{mín}\left\{g(z), \overline{c} + \sum_{m\in\mathcal{G}} q_m E_{\theta_0}\rho\left(z\frac{f_1^m\big(X_m^{(m)}\big)}{f_0^m\big(X_m^{(m)}\big)}\right)\right\}\right) \\
\Rightarrow \qquad & \\
\lambda &= \text{mín}\{\lambda_0, \overline{c} + \lambda\} \qquad (3.18)
\end{aligned}$$

y como $\overline{c} > 0$ entonces $\overline{c} + \lambda > \lambda$. De lo cual

$$\lambda = \lambda_0$$

Por lo tanto:

$$\lim_{z\to\infty} \rho(z) = \lambda_0$$

■

Definamos ahora la función h mediante:

$$h(z) = \sum_{m\in\mathcal{G}} q_m E_{\theta_0} \rho\left(z \frac{f_1^m\left(X_m^{(m)}\right)}{f_0^m\left(X_m^{(m)}\right)} \right) \tag{3.19}$$

Con base en el Lema 7, esta función está definida para toda $z \in [0,\infty)$ y tiene las mismas características de la función $\rho(z)$, a saber:

(1) $h(0) = 0$ y $\lim_{z\to\infty} h(z) = \lambda_0$;

(2) $h(z)$ es cóncava y continua en el intervalo $[0,\infty)$ y

(3) $h(z)$ es creciente en $[0,\infty)$.

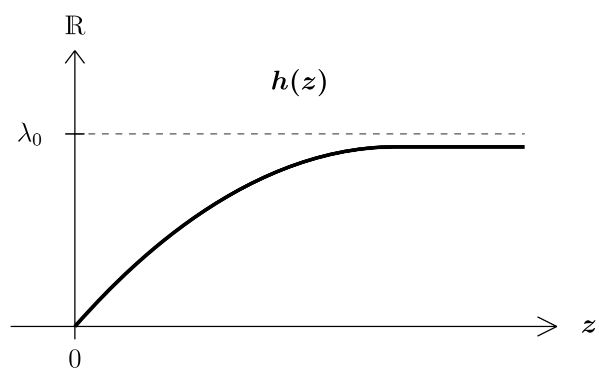

Figura 3.1: Función $h(z)$

Entonces (3.17) se puede reescribir como:

$$g(z) \quad \leq \quad \bar{c} + h(z) \tag{3.20}$$

Ahora trabajemos con la función $g(z) = \text{mín}\{\lambda_0, \lambda_1 z\}$, para esto sea:

$$\lambda = \frac{\lambda_0}{\lambda_1} \tag{3.21}$$

$g(z)$ es una función definida para todo $z \in [0, \infty)$ y con las siguientes características:

(1) $g(z)$ es continua, en $[0, \infty)$;

(2) $g(0) = 0 \quad$ y $\quad g(\lambda) = \lambda_0$;

(3) Si $z < \lambda$ entonces $\lambda_1 z < \lambda_0, \quad \Rightarrow \quad g(z) = \lambda_1 z$.

Por tanto, si $z \in [0, \lambda)$, $g(z)$ es una recta con pendiente positiva λ_1;

(4) Si $z \geq \lambda$ entonces $\lambda_0 \leq \lambda_1 z, \quad \Rightarrow \quad g(z) = \lambda_0$.

Por tanto, si $z \in [\lambda, \infty)$, $g(z)$ es una recta horizontal con valor λ_0;

(5) Finalmente: $\lim\limits_{z \to \infty} g(z) = \lambda_0$.

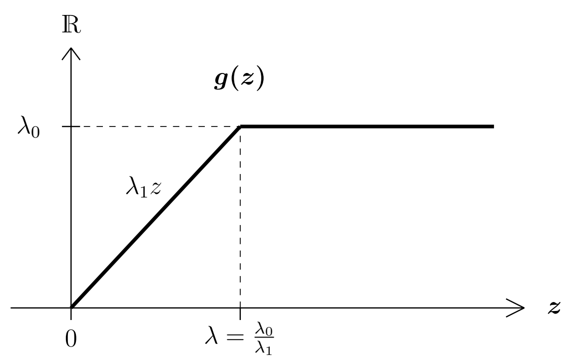

Figura 3.2: Función $g(z)$

Ahora vamos a analizar conjuntamente las funciones $g(z)$ y $h(z)$, para lo cual primero:

Lema 8.

$$\textit{Si } z \in [0,\infty) \textit{ entonces } h(z) \leq g(z).$$

Demostración:

Dada la desigualdad de Jensen $E\rho(X) \leq \rho\big(EX\big)$ con ρ cóncava, sea $z \in [0,\infty)$ fija, entonces:

$$\begin{aligned}
E_{\theta_0}\rho\left(z\frac{f_1^m\big(X_m^{(m)}\big)}{f_0^m\big(X_m^{(m)}\big)}\right) &\leq \rho\left(z\ E_{\theta_0}\ \frac{f_1^m\big(X_m^{(m)}\big)}{f_0^m\big(X_m^{(m)}\big)}\right)\\
&= \rho\left(z\sum_{x_m^{(m)}}\frac{f_1^m\big(x_m^{(m)}\big)}{f_0^m\big(x_m^{(m)}\big)}\Big(f_0^m\big(x_m^{(m)}\big)\Big)\right)\\
&= \rho\left(z\sum_{x_m^{(m)}} f_1^m\big(x_m^{(m)}\big)\right)\\
&= \rho\left(z\right)
\end{aligned}$$

Esta desigualdad se tiene para cualquier $z \in [0,\infty)$ fija entonces ahora:

$$\begin{aligned}
\sum_{m\in\mathcal{G}} q_m E_{\theta_0}\rho\left(z\frac{f_1^m\big(X_m^{(m)}\big)}{f_0^m\big(X_m^{(m)}\big)}\right) &\leq \sum_{m\in\mathcal{G}} q_m\rho\left(z\right)\\
&= \rho(z)\\
&\leq g(z).
\end{aligned}$$

$\therefore \quad h(z) \leq g(z)$ para toda $z \in [0,\infty)$. ■

Sea la función D donde:

$$D(z) \;=\; g(z) - h(z). \tag{3.22}$$

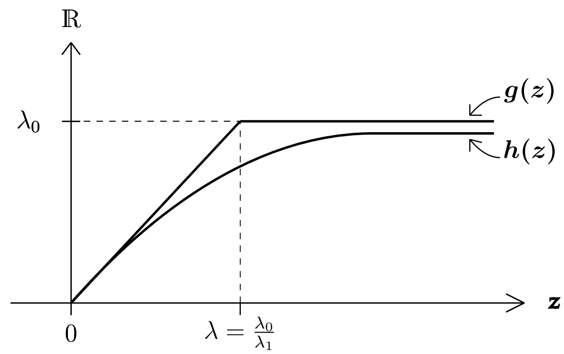

Figura 3.3:

Utilizando la definición de $g(z)$ se tiene que:

$$D(z) = \begin{cases} \lambda_1 z - h(z), & \text{si } z \leq \lambda; \\ \lambda_0 - h(z), & \text{si } z \geq \lambda. \end{cases}$$

Con base en las propiedades de g y h, la función D hereda las siguientes propiedades: (véase 3.4)

(1) $D(z)$ está definida para todo $z \in [0, \infty)$ y es continua para toda z;

(2) Por Lema 8, $D(z) \geq 0 \quad \forall z \in [0, \infty)$;

(3) $D(0) = 0$ y $\lim\limits_{z \to \infty} D(z) = 0$.

(4) En cada intervalo $[0, \lambda]$ y $[\lambda, \infty)$ por separado $D(z)$ es convexa.
Porque en cada intervalo se tiene la suma de dos funciones convexas, una lineal $g(z)$ y otra convexa $-h(z)$.

(5) Si $z < \lambda$ entonces $D(z)$ es creciente.

Debido a que, dados $0 \leq z_1 < z_2 < \lambda$, entonces $0 \leq \frac{z_1}{z_2} < 1$.

Se puede escribir a z_1 como una combinación de z_2 y 0.

$$z_1 = \frac{z_1}{z_2} z_2 + \left(1 - \frac{z_1}{z_2}\right) 0$$

Dado que $D(z)$ es convexa en el intervalo $[0, \lambda)$, entonces

$$\begin{aligned} D(z_1) &= D\left(\frac{z_1}{z_2} z_2 + \left(1 - \frac{z_1}{z_2}\right) 0\right) \\ &\leq \frac{z_1}{z_2} D(z_2) + \left(1 - \frac{z_1}{z_2}\right) D(0) \leq D(z_2) \end{aligned}$$

$\therefore \quad z_1 < z_2 \; \Rightarrow \; D(z_1) \leq D(z_2)$.

(6) Si $z \geq \lambda$ entonces $D(z)$ es decreciente.

Esto es porque, dados $\lambda \leq z_1 < z_2$ y como $h(z)$ es creciente, entonces:

$$\begin{aligned} & h(z_1) &\leq& \; h(z_2) \\ \Rightarrow \quad & -h(z_1) &\geq& \; -h(z_2) \\ \Rightarrow \quad & \lambda_0 - h(z_1) &\geq& \; \lambda_0 - h(z_2) \\ \Rightarrow \quad & D(z_1) &\geq& \; D(z_2) \end{aligned}$$

$\therefore \quad z_1 < z_2 \; \Rightarrow \; D(z_1) \geq D(z_2)$.

(7) Con base en (5) y (6), $D(z)$ tiene un valor máximo en el punto $z = \lambda$, *i. e.* $D(\lambda)$ es máximo. De lo cual, la imagen de $D(z)$ es $[0, D(\lambda)]$.

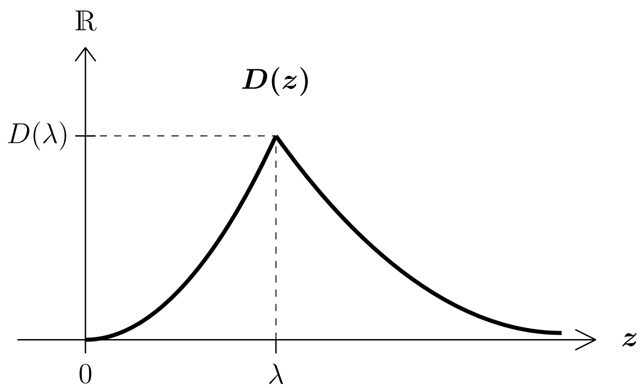

Figura 3.4: Función $D(z)$

Sabemos que $0 < \overline{c}$, definido en (1.17), es el costo promedio de un grupo, entonces éste valor comparado con el valor máximo de $D(z)$ tiene tres opciones, a saber: $\overline{c} < D(\lambda)$, $\overline{c} = D(\lambda)$ y $\overline{c} > D(\lambda)$ (véase gráficas 3.5, 3.1). Por el momento sólo nos enfocaremos en el primer caso, los otros dos casos serán analizados posteriormente.

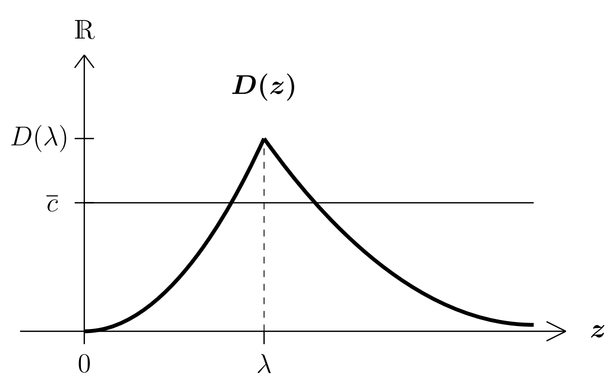

Figura 3.5: $\overline{c} < D(\lambda)$

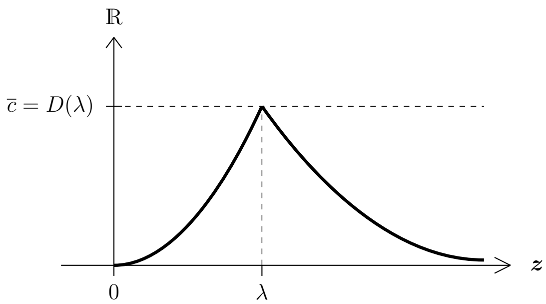

Figura 3.6: $\bar{c} = D(\lambda)$

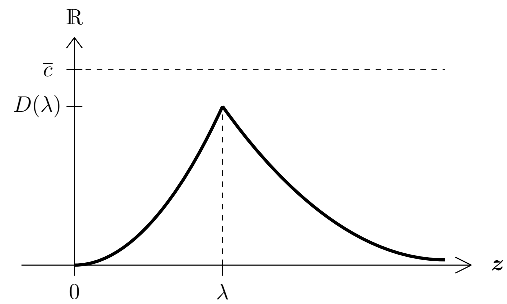

Figura 3.7: $\bar{c} > D(\lambda)$

Lema 9.

Si $0 < \bar{c} < D(\lambda)$ *entonces existen únicos* A *y* B *tales que:*

$$D(A) = \bar{c} \quad con \quad A \in (0, \lambda) \ y \ ; \tag{3.23}$$

$$D(B) = \bar{c} \quad con \quad B \in (\lambda, \infty). \tag{3.24}$$

Demostración

Sabemos que D es una función creciente en el intervalo $[0, \lambda)$, $D(\lambda)$ es máximo, y por hipótesis $0 < \bar{c} < D(\lambda)$ (véase figura 3.5). Entonces por continuidad de D existe un punto $A \in [0, \lambda)$ tal que $D(A) = \bar{c}$. Sin embargo $A \neq 0$ porque si $A = 0$ entonces $D(0) = 0 = \bar{c}$, contradicción porque $0 < \bar{c}$. Por tanto $A \in (0, \lambda)$.

Ahora se sabe que D es decreciente en el intervalo $[\lambda, \infty)$ y $\overline{c} < D(\lambda)$. También sabemos que $\lim\limits_{z\to\infty} D(z) = 0$, entonces existe $z \in [\lambda, \infty)$ tal que $D(z) < \overline{c}$. De aquí se tiene:

$$D(z) < \overline{c} < D(\lambda)$$

Entonces, por la continuidad de D se tiene que existe $B \in [\lambda, \infty)$ tal que $D(B) = \overline{c}$ y $\lambda < B < z < \infty$.

De lo anterior se concluye que existen $0 < A < \lambda < B < \infty$ tales que
$D(A) = D(B) = \overline{c}$.

Ahora demostraremos que dichos puntos A y B son únicos cada uno en su intervalo. En efecto, supongamos que existe otro $A' \in (0, \lambda)$ con $D(A') = \overline{c}$ y tal que $A' \neq A$. Supongamos que $A' < A$, Sabemos que $A' \neq 0$ y $A \neq 0$ entonces tomemos $0 < \frac{A'}{A} < 1$. Sea

$$A = \Big(\frac{A'}{A}\Big)A + \Big(1 - \frac{A'}{A}\Big)0.$$

Sabemos que $D(z)$ es convexa en el intervalo $[0, \lambda)$, entonces

$$\begin{aligned} D(A) &= D\bigg(\Big(\frac{A'}{A}\Big)A + \Big(1 - \frac{A'}{A}\Big)0\bigg) \\ &\leq \Big(\frac{A'}{A}\Big)D(A) + \Big(1 - \frac{A'}{A}\Big)D(0) \\ &\leq \Big(\frac{A'}{A}\Big)\overline{c}. \end{aligned}$$

Por tanto $\overline{c} \ \leq \dfrac{A'}{A}\ \overline{c}$ además $\overline{c} \neq 0$. Lo cual es una contradicción porque $0 < \frac{A'}{A} < 1$.
Entonces $A' = A$

Si $A < A'$, se procede análogamente intercambiando dichas constantes. Por lo tanto el punto A es único.

Ahora supongamos que existe otro punto $B' \in (\lambda, \infty)$ tal que $D(B') = \overline{c}$ y $B' \neq B$. Supongamos que $B' < B$ (si $B < B'$ se procede igual intercambiando B por B' y viceversa).

Se sabe que $\exists\ z > B$ tal que $D(z) < \overline{c}$ entonces se tiene $B' < B < z$ con base en esto tomemos las siguientes proporciones sean:

$$\alpha = \frac{z - B}{z - B'} \quad \Rightarrow \quad (1 - \alpha) = \frac{B - B'}{z - B'}$$

donde $0 < \alpha \leq 1$ entonces se puede escribir:

$$\begin{aligned} B &= \frac{z - B}{z - B'}B' + \frac{B - B'}{z - B'}z \\ \Rightarrow \quad D(B) &= D\left(\left(\frac{z - B}{z - B'}\right)B' + \left(\frac{B - B'}{z - B'}\right)z\right) \end{aligned}$$

Como D es convexa en el intervalo $[\lambda, \infty)$, entonces

$$\begin{aligned} D(B) &\leq \left(\frac{z - B}{z - B'}\right)D(B') + \left(\frac{B - B'}{z - B'}\right)D(z) \\ \Rightarrow \quad \overline{c} &\leq \left(\frac{z - B}{z - B'}\right)\overline{c} + \left(\frac{B - B'}{z - B'}\right)D(z) \\ \Rightarrow \quad \left(\frac{B - B'}{z - B'}\right)\overline{c} &\leq \left(\frac{B - B'}{z - B'}\right)D(z) \\ \Rightarrow \quad \overline{c} &\leq D(z). \end{aligned}$$

Lo cual es una contradicción porque $D(z) < \overline{c}$.

Entonces $B' = B$ y por lo tanto el punto B es único. ∎

En el lema 9 se demuestra la existencia de un único $A \in (0, \lambda)$ tal que:

$$\begin{aligned} \overline{c} &= D(A) = g(A) - h(A) \\ \Rightarrow \quad g(A) &= \overline{c} + h(A). \end{aligned} \tag{3.25}$$

Análogamente existe un único $B \in (\lambda, \infty)$ tal que:

$$g(B) \;=\; \overline{c} + h(B). \tag{3.26}$$

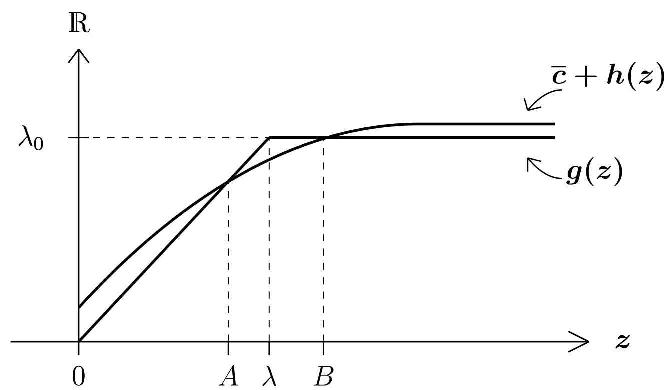

Figura 3.8:

Lema 10.

Si A *y* B *cumplen las ecuaciones* (3.25) *y* (3.26)*, entonces*

$$A < z < B \;\; \textit{si y solamente si} \;\; g(z) > \overline{c} + h(z). \tag{3.27}$$

Demostración:

Necesidad:

Si $z \in (A, \lambda]$ entonces $A < z \leq \lambda \quad \Rightarrow \quad D(A) < D(z)$ porque $D(z)$ es creciente y A es único.

De aquí se sigue que $\bar{c} < g(z) - h(z)$ y por tanto $\bar{c} + h(z) < g(z)$.

De la misma manera se tiene que si $z \in [\lambda, B)$ entonces $\lambda \leq z < B$, como D es decreciente entonces $D(B) \leq D(z)$ sin embargo ser B único entonces $D(B) < D(z)$. De lo cual $\bar{c} + h(z) < g(z)$.

Por tanto:

$$\text{si } \; z \in (A, B) \text{ entonces } \; \bar{c} + h(z) < g(z).$$

Suficiencia:

Por otro lado si $z \in [0, A]$ entonces podemos escribir a este punto de la forma $z = (1-a)A$ con $0 \leq a \leq 1$. Se sabe que $h(0) = 0$, entonces tomemos el segmento que va de 0 a z, como $h(z)$ es cóncava entonces:

$$h(z) = h\Big((1-a)A\Big) \geq (1-a)h(A) = (1-a)\Big(g(A) - \bar{c}\Big)$$

De lo cual

$$\begin{aligned} \bar{c} + h(z) &\geq \bar{c} + (1-a)\Big(g(A) - \bar{c}\Big) \\ &= a\bar{c} + (1-a)g(A) \\ &\geq (1-a)g(A) = g\Big((1-a)A\Big) = g(z) \end{aligned}$$

$$\therefore \qquad \text{si } \; z \in [0, A] \text{ entonces } \; \bar{c} + h(z) \geq g(z).$$

Sea ahora $z \in [B, \infty)$, entonces $B \leq z$. Como h es creciente y cóncava, entonces

$$h(B) \leq h(z) \quad \Rightarrow \quad 0 \leq h(z) - h(B) \quad \Rightarrow \quad \lambda_0 \leq h(z) - h(B) + \lambda_0$$

Por propiedades de la función g sabemos que $g(B) = \lambda_0 = g(z)$, ya que $\lambda < B \leq z$.

Entonces sustituyendo en la ecuación anterior:

$$\begin{aligned} \underbrace{\lambda_0}_{g(z)} &\leq h(z) - h(B) + \underbrace{\lambda_0}_{g(B)} \\ \Rightarrow & \\ g(z) &\leq h(z) + g(B) - h(B) \\ \Rightarrow & \\ g(z) &\leq h(z) + \bar{c}. \end{aligned}$$

$$\therefore \qquad \text{si } \; z \in [B, \infty) \text{ entonces } \; \bar{c} + h(z) \geq g(z).$$

Por lo tanto, si $z \notin (A, B)$ entonces $g(z) \ngtr \bar{c} + h(z)$. ■

Con base en el lema anterior 10, se obtiene su equivalente:

$$z \notin (A, B) \qquad \Leftrightarrow \qquad g(z) \leq \bar{c} + h(z) \tag{3.28}$$

con ciertas constantes A, B tales que $0 < A < B < \infty$ y, respectivamente, cumplen las ecuaciones (3.25) y (3.26). No olvidemos que dichas constantes A y B existen bajo la hipótesis de que $h(\lambda) < \lambda_0 - \bar{c}$. Por lo cual la expresión (3.16) se reescribe como:

$$\psi^*_{\boldsymbol{n}} = I_{\left\{Z_{\boldsymbol{n}} \notin (A,B)\right\}} \tag{3.29}$$

Reescribiendo ahora la regla de decisión ϕ^* se tiene:

$$\begin{aligned} \phi^*_{\boldsymbol{n}} &= I_{\{\lambda_0 f_0^{\boldsymbol{n}} \leq \lambda_1 f_1^{\boldsymbol{n}}\}} \\ &= I_{\left\{\frac{\lambda_0}{\lambda_1} \leq \frac{f_1^{\boldsymbol{n}}}{f_0^{\boldsymbol{n}}}\right\}} \\ &= I_{\left\{\frac{\lambda_0}{\lambda_1} \leq Z_{\boldsymbol{n}}\right\}} \end{aligned}$$

Como para aplicar la regla de decisión ϕ^*, se necesita primero haber aplicado la regla de paro ψ^*, y si el hecho que define a dicha regla ya sucedió $\big(Z_{\boldsymbol{n}} \notin (A,B)\big)$, entonces la regla de decisión se convierte en:

$$\phi^*_{\boldsymbol{n}} \quad = \quad I_{\{B \leq Z_{\boldsymbol{n}}\}} \tag{3.30}$$

Entonces como $A < \frac{\lambda_0}{\lambda_1} < B$, resulta que la prueba secuencial óptima termina en el momento en el que $Z_{\boldsymbol{n}} \notin (A,B)$.

Hasta este momento hemos reescrito las componentes (2.75) de la prueba (ψ^*, ϕ^*) en términos de la razón de probabilidades $Z_{\boldsymbol{n}}$ como:

$$\begin{aligned} \psi^*_{\boldsymbol{n}} &= I_{\{Z_{\boldsymbol{n}} \notin (A,B)\}}, \\ \phi^*_{\boldsymbol{n}} &= I_{\{Z_{\boldsymbol{n}} \geq B\}}. \end{aligned} \tag{3.31}$$

para ciertas constantes A, B tales que $0 < A < \frac{\lambda_0}{\lambda_1} < B < \infty$ y bajo el hecho de que $h(\lambda) < \lambda_0 - \overline{c}$. Rechazando H_0 si $Z_{\boldsymbol{n}} \geq B$ y aceptándola si $Z_{\boldsymbol{n}} \leq A$. Donde las constantes A y B se determinan a través de las ecuaciones:

$$g(A) \quad = \quad \overline{c} + \sum_{m \in \mathcal{G}} q_m E_{\theta_0} \rho \left(A \frac{f_1^m\big(X_m^{(m)}\big)}{f_0^m\big(X_m^{(m)}\big)} \right) \tag{3.32}$$

$$g(B) \quad = \quad \overline{c} + \sum_{m \in \mathcal{G}} q_m E_{\theta_0} \rho \left(B \frac{f_1^m\big(X_m^{(m)}\big)}{f_0^m\big(X_m^{(m)}\big)} \right). \tag{3.33}$$

Teorema 6.

Dada la prueba secuencial (ψ^*, ϕ^*) *cuyas respectivas componentes están definidas en* (3.31), *con* A *y* B *definidas en* (3.32) *y* (3.33). *Entonces para cualquier prueba* (ψ, ϕ) *tal que*

$$\alpha(\psi, \phi) \leq \alpha(\psi^*, \phi^*), \quad \beta(\psi, \phi) \leq \alpha(\psi^*, \phi^*)$$

Se cumple:

$$K(\psi) \geq K(\psi^*).$$

Demostración

En el teorema 5 se demostró que la prueba secuencial (ψ^*, ϕ^*) es óptima en el sentido de que minimiza el costo del experimento. Donde las componentes de la prueba están determinadas por las ecuaciones (2.75). Pero dichas ecuaciones, en esta sección, fueron reescritas por las relaciones en (3.31) en las cuales se tienen ciertas constantes A y B, que son el resultado de laas ecuaciones (3.32) y (3.33). Aunque cabe mencionar que la existencia de dichas constantes está sujeto a la condición $0 < \bar{c} < D(\lambda)$ o su equivalente $h(\lambda_0) < \lambda_0 - \bar{c}$. Por tanto, para esta nueva reescritura la prueba (ψ^*, ϕ^*) es óptima.

La prueba secuencial (ψ^*, ϕ^*) con sus correspondientes componentes previamente definidas, es un caso particular de la prueba secuencial conocida como *Prueba Secuencial Aleatoria de la Razón de Probabilidades o (RSPRT)*. La prueba *RSPRT* será trabajada con más detalle en la subsiguiente. La prueba (ψ^*, ϕ^*) es particular porque ella óptima ciertas constantes dadas [1], a diferencia de la RSPRT que es óptima para cualesquiera constantes.

Finalmente para esta prueba (ψ^*, ϕ^*), caso particular de la *RSPRT*, tomemos su tiempo de paro asociado. Sea

$$\tau^{\psi^*} = \text{mín}\Big\{k : Z_{\boldsymbol{\nu}} \notin (A, B)\Big\} \tag{3.34}$$

Este tiempo de paro tiene el mismo sentido práctico que el trabajado en [10] (página 434). Dicho tiempo está directamente asociado a la regla de paro ψ^*, debido a que cuando ella para $\big(Z_{\boldsymbol{n}} \notin (A, B)\big)$ se tendrá un tiempo de paro óptimo [2].

3.2. Una prueba secuencial trivial

Todo lo anterior se trabaja cuando se tiene un experimento el cual por lo menos tiene una observación (o dato), porque con base en dicho dato se comienza el análisis secuencial (regla de paro, regla de decisión, etcétera). Sin embargo todo el análisis anterior podría conllevar a mejores resultados ¡sin observar ningún dato!

Existe una prueba secuencial que no necesita de datos para ser realizada, esta prueba es conocida como la *Prueba Secuencial Trivial*, en esta sección analizaremos a esta pareja.

Sea (ψ_0, ϕ_0) la prueba secuencial trivial, la cual no necesita de ningún dato para aplicarse. Debido a ello la "función de paro" es siempre uno $(\psi_0 \equiv 1)$, ya que el experimento

[1] Cuya estructura ha sido determinada.

[2] O un mínimo de grupos observados en el experimento.

se detiene antes del primer paso, cuando no hay datos. Analicemos ahora la función de decisión ϕ_0, para esto tómese la función de Lagrange aplicada a la prueba trivial, entonces:

$$L(\psi_0, \phi_0) = K(\psi_0) + \lambda_0 \alpha(\psi_0, \phi_0) + \lambda_1 \beta(\psi_0, \phi_0). \tag{3.35}$$

Como no hay observaciones en el experimento entonces el costo de éste será nulo $\big(K(\psi_0) = 0\big)$, de lo cual

$$L(\psi_0, \phi_0) = \lambda_0 \alpha(\psi_0, \phi_0) + \lambda_1 \beta(\psi_0, \phi_0) \tag{3.36}$$

Esta ecuación solamente trabaja con las probabilidades de los errores tipo I y tipo II, y con sus respectivos pesos (λ_0, λ_1) en la prueba.

Si la decisión final es aceptar la hipótesis nula H_0 $\big(\phi_0 \equiv 0\big)$, entonces la probabilidad de rechazar H_0 siendo ésta verdadera es nula, esto es $\alpha(\psi_0, \phi_0) = 0$. Y la probabilidad de aceptar H_0, siendo ésta falsa, vale uno, de lo cual $\beta(\psi_0, \phi_0) = 1$. Por tanto, la función de Lagrange resulta:

$$\lambda_0 \alpha(\psi_0, \phi_0) + \lambda_1 \beta(\psi_0, \phi_0) = \lambda_1 \tag{3.37}$$

Análogamente, si la decisión final es rechazar la hipótesis nula H_0 $\big(\phi_0 \equiv 1\big)$, entonces la probabilidad de rechazar H_0, siendo ésta verdadera, es uno $\big(\alpha(\psi_0, \phi_0) = 1\big)$, y $\beta(\psi_0, \phi_0) = 0$ y por tanto:

$$\lambda_0 \alpha(\psi_0, \phi_0) + \lambda_1 \beta(\psi_0, \phi_0) = \lambda_0 \tag{3.38}$$

En resumen

$$\alpha(\psi_0, \phi_0) = \begin{cases} 1, & \text{si } \lambda_0 \leq \lambda_1; \\ 0, & \text{si } \lambda_0 > \lambda_1. \end{cases} \qquad \beta(\psi_0, \phi_0) = \begin{cases} 0, & \text{si } \lambda_0 \leq \lambda_1; \\ 1, & \text{si } \lambda_0 > \lambda_1. \end{cases}$$

Con base en (3.37) y (3.38) se sigue que:

$$\begin{aligned} \min_{\phi_0} L(\psi_0, \phi_0) &= \min\{\lambda_0, \lambda_1\} \\ &= \lambda_0 \alpha(\psi_0, \phi_0) + \lambda_1 \beta(\psi_0, \phi_0) \end{aligned} \tag{3.39}$$

De donde la regla de decisión resulta:

$$\phi_0 = I_{\{\lambda_0 \leq \lambda_1\}}$$

Utilizando la definición (3.3) de la función g se tiene que

$$g(1) = L(\psi_0, \phi_0) \tag{3.40}$$

Este valor de la función de Lagrange $\big(L(\psi_0, \phi_0)\big)$ es mínimo, debido a que no hay datos, comparado con el valor mínimo de la función de Lagrange $\big(L(\psi^*, \phi^*)\big)$ cuando se tiene por lo menos una observación, y este último valor ya lo calculamos en el proceso anterior $\big($teorema 4 (2.64), tomando en cuenta que $L(\psi^*) = L(\psi^*, \phi^*)\big)$. De lo cual:

$$\begin{aligned} L(\psi^*, \phi^*) &= \bar{c} + \sum_{n \in \mathcal{G}} q_n \sum_{x^n} V_n \\ &= \bar{c} + \sum_{n \in \mathcal{G}} q_n \sum_{x^n} U_n(Z_n) f_0^n \\ &= \bar{c} + \sum_{n \in \mathcal{G}} q_n \sum_{x^n} f_0^n \rho(Z_n) \\ &= \bar{c} + h(1) \\ &\geq L(\psi_0, \phi_0) = g(1) \end{aligned}$$

Esto es, $g(1) \leq \bar{c} + h(1)$. Ya que la función de Lagrange aplicada a la prueba secuencial trivial cuando no se tienen datos es menor a la función de Lagrange aplicada a la prueba óptima cuando se tiene al menos un dato. Entonces utilizando la definición de la función h en la ecuación anterior, se tiene que:

$$g(1) \leq \bar{c} + \sum_{m \in \mathcal{G}} q_m E_{\theta_0} \rho \left(\frac{f_1^m \big(X_m^{(m)}\big)}{f_0^m \big(X_m^{(m)}\big)} \right) \tag{3.41}$$

Con base en el lema 10 (3.27) la desigualdad anterior significa que $1 \notin (A, B)$. Entonces una prueba secuencial que conlleva también a un costo mínimo y no necesita de observaciones para ser realizada, es la prueba secuencial trivial (ψ_0, ϕ_0) donde:

$$\begin{aligned} \psi_0 &\equiv 1, \\ \phi_0 &= I_{\{\lambda_0 \leq \lambda_1\}}. \end{aligned} \tag{3.42}$$

Para NO encontrarnos en el caso de trabajar con dicha prueba trivial (ψ_0, ϕ_0) es necesario que se tenga la condición $1 \notin (A, B)$. Así que solamente en el caso $A < 1 < B$ la prueba secuencial óptima tiene un sentido práctico, ya que en caso contrario con la prueba trivial se obtienen mejores resultados.
Los casos "límite" cuando $1 = A < B$ o $A < B = 1$ pueden ser incluidos en consideración, por lo menos para fines teóricos, ya que en este caso la prueba trivial da exactamente lo mismo que la RSPRT y no algo mejor.

Recordemos que para la existencia de las contantes A y B éstas están sujetas a la condición $0 < \overline{c} < D(\lambda)$ (Lema 9). En los casos en los que dicha condición no se cumpla, *i. e.* cuando $\overline{c} \geq D(\lambda)$, la función $g(z)$ es siempre menor (o igual) a la función $\overline{c} + h(z)$, y se tendrá que para toda $z \in [0, \infty)$, $g(z) \leq \overline{c} + h(z)$ (véanse figuras 3.6, 3.7). Entonces, en particular $g(1) < \overline{c} + h(1)$, lo cual significa que la prueba secuencial trivial cuando no se toman observaciones es mejor comparada con la prueba que almenos toma una observación. Por tanto, la prueba secuencial óptima que se trabajará será la prueba secuencial trivial (ψ_0, ϕ_0), ya que ésta nos lleva a mejores resultados en estos casos.

Hasta este momento hemos demostrado por completo que la prueba secuencial (ψ^*, ϕ^*) es óptima, con unas contantes A y B especificas. En lo subsiguiente se demostrará la optimalidad de la *RSPRT* para cualesquiera constantes A y B.

3.3. Optimalidad de la RSPRT

A partir de nuestra prueba óptima (ψ^*, ϕ^*) se obtienen las contantes particulares A y B que resultan de las ecuaciones (3.32) y (3.33). Una pregunta ahora sería, ¿el recíproco también se sigue?, esto es ¿para cualesquiera constantes A y B, la RSPRT será óptima? Aunque esta pregunta es verosímil porque que para *cualesquiera* constantes A y B tales que $0 < A < B < \infty$ pueden existir otras constantes $\bar{c}$, λ_0, y λ_1 con las cuales $L(\psi)$ es minimizada por la prueba RSPRT, ya que sólo hay dos restricciones por cumplir (3.32), (3.33) y las variables son tres: $\bar{c}$, λ_0 y λ_1. El teorema que formaliza lo anterior es el siguiente:

Teorema 7.

Sean A, B dos constantes tales que $0 < A < B < \infty$ y sea (ψ^*, ϕ^*) una prueba RSPRT, cuyas componentes están definidas en (3.31)*. Entonces existen $\lambda_0 > 0$, $\lambda_1 > 0$ tales que para cualquier prueba secuencial (ψ, ϕ) se tiene:*

$$K(\psi) + \lambda_0\alpha(\psi, \phi) + \lambda_1\beta(\psi, \phi) \geq K(\psi^*) + \lambda_0\alpha(\psi^*, \phi^*) + \lambda_1\beta(\psi^*, \phi^*) \tag{3.43}$$

Observación: Nótese que este teorema establece la optimalidad de la prueba (ψ^*, ϕ^*) en el mismo sentido del teorema 4, salvo que en el primero se parte de dos constantes cualesquiera. En el teorema 4 se trabaja con $L(\psi^*)$ que contiene implícitamente a ϕ^* de la siguiente manera:

$$L(\psi, \phi^*) = L(\psi) \geq L(\psi^*) = L(\psi^*, \phi^*)$$

La demostración de la existencia de λ_0 y λ_1 tales que conllevan a la optimalidad de la prueba (ψ^*, ϕ^*) puede ser revisada en [12] páginas 309-313. Dicha demostración es un resultado clásico en el campo del análisis secuencial, ésta fue publicada en el trabajo de Lehmann en 1959 [7]. Este resultado, en dicho trabajo, no es muy entendible, incluso algunos otros autores cuando la utilizan hacen referencia directa al trabajo de Lehmann, sin volver a escribir la demostración. Cabe mencionar que tal demostración tiene un estilo Bayesiano, mas en el trabajo [12] el estilo es más acorde al nuestro. Por la extensión de la demostración y la semejanza con nuestro trabajo ella será sólo referida debido a que en el trabajo citado sí se presenta la demostración completa, y es entendible. Por lo cual nos es pertinente sólo hacer la referencia.

La prueba (ψ^*, ϕ^*) del teorema anterior es conocida como ***la Prueba Secuencial Aleatoria de la Razón de Probabilidades*** o *RSPRT* (por sus siglas en inglés, véase [10]). Se ha demostrado que la *RSPRT* en un experimento secuencial conlleva a una función de Lagrange minima y por tal a un costo mínimo. Por tanto ella es óptima. El siguiente teorema establece la optimalidad con respecto al costo bajo cualquier parámetro (θ_0 o θ_1).

Teorema 8.

Sean A, B *dos constantes tales que* $0 < A < B < \infty$ *y sea* (ψ^*, ϕ^*) *una prueba RSPRT, cuyas componentes están dadas por* (3.31). *Entonces para cualquier prueba secuencial* (ψ, ϕ) *tal que:*

$$\alpha(\psi, \phi) \leq \alpha(\psi^*, \phi^*) \quad y \quad \beta(\psi, \phi) \leq \beta(\psi^*, \phi^*) \tag{3.44}$$

se tiene:

$$K(\theta_0; \psi) \geq K(\theta_0; \psi^*) \quad y \quad K(\theta_1; \psi) \geq K(\theta_1; \psi^*). \tag{3.45}$$

la desigualdades en (3.45) *son estrictas si es estricta, por lo menos, una de las desigualdades de* (3.44).

Demostración:

Puede observarse que la primera desigualdad de (3.45) $\big(K(\theta_0; \psi) \geq K(\theta_0; \psi^*)\big)$ es consecuencia del teorema 7 (3.43) y del teorema 1. Porque el primero establece que la prueba (ψ^*, ϕ^*) minimiza la función de Lagrange para cualesquiera constantes A y B, y el segundo establece que cualquier prueba que tenga una función de Lagrange menor con las cotas establecidas en (3.44) conlleva a un costo mínimo. Por lo tanto la prueba (ψ^*, ϕ^*) minimiza el costo bajo del experimento bajo el parámetro θ_0.

Ahora sólo nos concretaremos a la demostración de $K(\theta_1; \psi) \geq K(\theta_1; \psi^*)$, para cualquier prueba secuencial (ψ, ϕ) con $\alpha(\psi, \phi) \leq \alpha(\psi^*, \phi^*)$ y $\beta(\psi, \phi) \leq \beta(\psi^*, \phi^*)$.

La demostración de este hecho es hacer exactamente todo el procedimiento anterior hasta la demostración del teorema 6, salvo por la parte que el parámetro utilizado en todo el proceso fue θ_0 y ahora el parámetro a utilizar es θ_1. Como dichos parámetros son intercambiables [3] entonces podemos decir que el anterior procedimiento puede volver a ser aplicado, conllevando a los mismo resultados. Por tanto se cumplirá la segunda desigualdad de (3.45). A continuación sólo marcaremos los puntos importantes para los cuales se trabajará con el parámetro θ_1.

1. El experimento es el mismo, con las mismas observaciones, tamaños de grupos, y sus respectivas variables aleatorias $X_1^{(\nu_1)}, X_2^{(\nu_2)}, \ldots, X_k^{(\nu_k)}, \ldots$.

2. Se tiene una prueba de hipótesis [4] donde:

$$H_0' : \theta = \theta_1 \quad vs. \quad H_1' : \theta = \theta_0.$$

3. Las reglas de paro ψ se definen igual que en la definición 1 y tienen la misma aplicación: para $k \geq 0$ con $\boldsymbol{n} \in \mathcal{G}^k$,

   - Si $\psi_{\boldsymbol{n}} = 0$, continua;
   - Si $\psi_{\boldsymbol{n}} = 1$, para.

   Como ψ es la misma, las función que dependen de ψ también se definen igual tal son los casos de τ^ψ, $s_{\boldsymbol{n}}^\psi$ y $t_{\boldsymbol{n}}^\psi$.

4. Por su parte, la regla de decisión ϕ sí tiene la misma definición, mas una interpretación diferente. A saber:

[3] Bajo sus respectivos ajustes.

[4] Sólo se hace un intercambio de parámetros.

- Si $\phi_{\boldsymbol{n}} = 0$, se acepta H_0' (θ_1);
- Si $\phi_{\boldsymbol{n}} = 1$, a favor de H_1' (θ_0).

Para relacionar estas nuevas funciones de decisión ϕ con las anteriores se tomará en este contexto una regla de decisión $1-\phi$ donde $1-\phi_{\boldsymbol{n}}$ tienen la misma interpretación del caso anterior.

5. Entonces, la definición de prueba secuencial en este contexto será la pareja $(\psi, 1-\phi)$. La cual resulta la **misma** que la establecida en la definición 5.

6. La probabilidad de que el tiempo de paro (τ^{ψ}) se detenga en un tiempo k dada en la ecuación (1.6) trabajaba con el parámetro θ libre, y no dependía de la regla de decisión, por tal dicha ecuación se trabaja ahora con θ_1.

7. Análogamente, la clase Ψ_{θ} se puede trabajar con el parámetro θ_1. Se tendrá la clase Ψ_{θ_1}.

8. También cambian las definiciones de las probabilidades de error tipo I y II. En este contexto se tiene:

$$\alpha'(\psi, 1-\phi) = P_{\theta_1}\big(\text{rechazar } H_0'\big) \quad \text{y} \quad \beta'(\psi, 1-\phi) = P_{\theta_0}\big(\text{aceptar } H_0'\big) \tag{3.46}$$

Se puede deducir fácilmente que

$$\alpha'(\psi, 1-\phi) = \beta(\psi, \phi) \quad \text{y} \quad \beta'(\psi, 1-\phi) = \alpha(\psi, \phi).$$

9. El costo del experimento también tenía el parámetro θ libre y no depende de la regla de decisión, por tal se tiene la misma relación dada en (1.22), pero ahora con θ_1. Esto es:

$$K(\theta_1; \psi) = \sum_{k=1}^{\infty} \sum_{\boldsymbol{n} \in \mathcal{G}^k} q_{\boldsymbol{n}}\ c(\boldsymbol{n})\ E_{\theta_1} s_{\boldsymbol{n}}^{\psi} + \ E_{\theta_1}\Big(c(\boldsymbol{\nu}^{\infty}) I_{\{\tau^{\psi}=\infty\}}\Big).$$

10. Ahora, el problema de optimización bajo la función de Lagrange es:

$$\begin{aligned} L(\psi; 1-\phi) &= K(\theta_1;\psi) + \lambda_0'\alpha'(\psi, 1-\phi) + \lambda_1'\beta'(\psi, 1-\phi). \\ &= K(\theta_1;\psi) + \lambda_0'\beta(\psi,\phi) + \lambda_1'\alpha(\psi,\phi). \end{aligned}$$

Obsérvese que esta función es semejante a nuestra función de Lagrange anterior, salvo por el costo y el orden de los pesos en las probabilidades de error.

11. Por tanto, todos los teoremas y lemas del capítulo 2 se aplican nuevamente a las nuevas relaciones, con sus respectivos ajustes.

12. En este nuevo contexto, la razón aleatoria de probabilidades se toma:

$$Z_{\boldsymbol{n}}' = \frac{f_0^{\boldsymbol{n}}}{f_1^{\boldsymbol{n}}}.$$

13. Se tiene las constantes A y B que resultan de la ecuaciones (3.32) y (3.33), con base en éstas se determinan nuevas constantes $A' = \frac{1}{B}$ y $B' = \frac{1}{A}$ obteniendo la condición de paro como $Z_{\boldsymbol{n}}' \notin (A', B')$ que equivale a $Z_{\boldsymbol{n}}' \notin \left(B^{-1}, A^{-1}\right)$.

14. Por lo cual la regla de paro que conlleva a un costo óptimo será aquella cuyas componentes tengan la estructura:

$$\psi_{\boldsymbol{n}}^* = I_{\left\{Z_{\boldsymbol{n}}' \notin (B^{-1}, A^{-1})\right\}} = I_{\left\{Z_{\boldsymbol{n}} \notin (A,B)\right\}}$$

la cual es exactamente la misma a la dada en (3.29).

15. Sin embargo para nuestro caso la regla de decisión tendrá la estructura:

$$1 - \phi_{\boldsymbol{n}}^* = 1 - I_{\left\{Z_{\boldsymbol{n}}' \geq A^{-1}\right\}} = I_{\left\{Z_{\boldsymbol{n}}' < A^{-1}\right\}} = I_{\left\{Z_{\boldsymbol{n}} > A\right\}}$$

Sin embargo recordemos que para aplicar la regla de decisión, previamente se ha cumplido la regla de paro. Por lo cual se tiene que $Z_{\boldsymbol{n}} \notin (A, B)$ y $Z_{\boldsymbol{n}} > A$, de donde el evento anterior es equivalente a $Z_{\boldsymbol{n}} \geq B$. Entonces

$$1 - \phi^*_{\boldsymbol{n}} = I_{\left\{Z_{\boldsymbol{n}} \geq B\right\}}$$

la cual es exactamente la regla de decisión definida en (3.30).

16. Por último, nuevamente se aplican los teoremas 8 y 1 a nuestro contexto y por lo tanto la prueba óptima cuyas componentes están definidas en (3.31) conlleva también a que $K(\theta_1; \psi) \geq K(\theta_1; \psi^*)$, para cualquier prueba secuencial (ψ, ϕ) con $\alpha(\psi, \phi) \leq \alpha(\psi^*, \phi^*)$ y $\beta(\psi, \phi) \leq \beta(\psi^*, \phi^*)$.

■

El teorema anterior ha demostrado que la prueba *RSPRT* es óptima en el sentido de que bajo su estructura se minimiza el costo de un experimento con observaciones independientes las cuales vienen en grupos de tamaño aleatorio. El costo es mínimo con cualquier parámetro establecido en la prueba de hipótesis simple.

CAPÍTULO 4

CONCLUSIONES Y PERSPECTIVAS.

Durante la segunda guerra mundial [1] alrededor de 1940, Abraham Wald desarrolló un método diferente para problemas de pruebas de hipótesis simples, con observaciones independientes [20]. En su trabajo, Wald utilizó *la prueba secuencial de la razón de probabilidades (SPRT)* para demostrar que el número promedio de observaciones en un experimento es más pequeño comparado con cualquier otro método que use un tamaño de muestra fijo. Con base en dicho trabajo se ha desarrollado el campo de la estadística conocido *análisis secuencial*. En este campo se han variado y creado nuevas hipótesis, condiciones, y elementos del trabajo de Wald para la optimización de pruebas.

En el año 2008 Nitis Mukhopadhyay y Basil de Silva [10] presentaron aplicaciones en las cuales se tenían planes de muestreo aleatorizados, debido a que se tenían grupos de observaciones de tamaño aleatorio. En este trabajo, expusieron la prueba conocida como

[1] 1939-1945.

RSPRT (modificación de la SPRT) y dejaron preguntas abiertas sobre las propiedades (*e. g. optimalidad*) de dicha prueba.

La investigación previa nos dio contexto para nuestra exploración, por el hecho de tomar en cuenta grupos de observaciones cuyo tamaño fuera aleatorio. Nuestro trabajo encontró un procedimiento y demostró la optimalidad de éste en pruebas secuenciales con grupos de tamaño aleatorio. Este proceso puede ser aplicado tanto a un contexto truncado (con un número fijo de etapas) como al caso no truncado; entonces se encontró un procedimiento *óptimo* para ambos casos. Por tanto, el resultado principal de esta investigación es **haber encontrado explícitamente *un procedimiento óptimo*** para una distribución dada, cuando los tamaños de los grupos no dependen de las observaciones previas y cuando la distribución de estos es fija. Con este procedimiento óptimo se ha obtenido una base teórica: para el desarrollo, salvo condiciones, de otros problemas tratados previamente ([4], [10], [15]) en el análisis secuencial; y para la extensión de la teoría de los planes de muestreo aleatorizados óptimos.

Otro importante resultado, obtenido en nuestro trabajo, ha sido la demostración de que *la prueba secuencial aleatoria de la Razón de Probabilidades* (RSPRT), propuesta en el trabajo Mukhopadhyay *et al* [10], es óptima [2]. En el sentido de que, ella minimiza el costo promedio de un experimento secuencial, en el que se tienen grupos de tamaño aleatorio. Sin embargo, es importante señalar que para nuestro caso, las variables asociadas a los tamaños de los grupos, se suponen variables aleatorias discretas, que toman un número finito de valores y son idénticamente distribuidas, a diferencia de la investigación anterior [*Íbid*], en la cual no se tienen estos supuestos. Cabe mencionar, como otra de

[2]Bajo hipótesis diferentes.

las contribuciones de la presente investigación, que el caso propuesto en la investigación, mencionada, con respecto a las distribuciones de los tamaños de los grupos, no conllevará a un resultado óptimo. Esto es, si se supone que los tamaños de los grupos tienen diferente distribución entonces la prueba $RSPRT$ **no** conllevará a un costo mínimo; nuestro trabajo da evidencia de este hecho.

Como futuras investigaciones se pueden analizar los mismos resultados con condiciones más generales como es el caso de suponer las observaciones como variables aleatorias con distribuciones continuas. Más aún, ya se tiene una base teórica para comenzar a trabajar el caso cuando las variables de los tamaños de los grupos no sean idénticamente distribuidos. Es decir, se puede ahora buscar un plan de muestreo aleatorizado óptimo.
También, como perspectiva, se tiene la búsqueda de planes de muestreo óptimos en el problema planteado por Cressie y Morgan [4], y Schmitz [15] sin que los tamaños de los grupos dependan de las observaciones.

Otro punto a señalar de esta tesis es que el caso demostrado por el pionero del campo Abraham Wald en el cual se toman observaciones de una en una, es un caso particular de nuestro esquema en donde todos los grupos son de tamaño uno. Así mismo en alguna parte del desarrollo de la investigación se tuvieron resultados generales que no necesitaban de la independencia de las observaciones, y estos procesos pudieran ayudar a investigaciones futuras.

El desarrollo teórico que le dimos al problema de optimización objeto de este trabajo, es parecido a los pasos principales del desarrollo clásico para problemas de paro óptimo (véase [3] y [16]). Sin embargo, no queda claro si la teoría clásica se pueda aplicar directamente a nuestro caso, donde las observaciones vienen en grupos de tamaño aleatorio.

Desde este punto de vista parece importante que los problemas de paro óptimo que vimos en esta tesis pueden ser resueltos usando herramientas bastante básicas de la teoría de probabilidad. Es muy probable que se puede extender los resultados aquí obtenidos a problemas más generales con grupos de tamaño aleatorio, manteniendo la misma estructura del desarrollo.

Finalmente sería muy valioso poder aplicar este nuevo conocimiento a la resolución de problemas prácticos en concreto. Desafortunadamente en nuestro país, los métodos del análisis secuencial no son aún aplicados a pesar de la optimalidad de sus resultados. A diferencia de otros países para los cuales dichos métodos ayudan directamente a la solución de problemas en contextos como: análisis clínicos, control de calidad, técnicas de confiabilidad, pruebas de educación, economía, finanzas, psicología, entre otros.

BIBLIOGRAFÍA

[1] Anderson, T. W., *A modification of the sequential probability ratio test to reduce the sample size*, The annals of Matheamtical Statistics, (1959): 165-197.

[2] Baum, C. W. y Veeravalli, V. V., *A sequential procedure for multhypothesis testing*, IEEE Transactions on Information Theory, (1994): 1994-2007.

[3] Chow, Y. S., Robbins, H., y Siegmund D, *Great Expectations: The Theory of Optimal Stopping*, Boston, Houghton Mifflin Company, 1971.

[4] Cressie, N. y Morgan, P. B., *The VRPT: A sequential testing procedure dominating the SPRT*. Econometric Theory 9 (1993): 431-450.

[5] Denne, J. S., y Jennison, C., *A group sequential test with updating of sample size*. Biometrika (2000): 125-134.

[6] Ghosh, M., Mukhopadhyay, N., and Sen, P., *Sequential Estimation*, Wiley series in probability and statistics. Probability and Statistics, 1997.

[7] Lehmann, E., *Testing Statistical Hypotheses*, Jonh Wiley & Sons, New York; Chapman & Hall, London, 1959.

[8] Lorden, G., *Structure of sequential tests minimizing an expected sample size*, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Geviete, 51 (1980): 291-302.

[9] Martinsek, Adam T., *The contributions of Robert A. Wijsman to sequential analysis*, Sequential Analysis, 25 (2006), 3-18.

[10] Mukhopadhyay, N. y De Silva B.M., *Theory and applications of a new methodology for the random sequential probability ratio test*, Statistical Methodology, 5 (2008): 424-453.

[11] Novikov, A., *Optimal sequential tests for two simple hypothesis*, Sequential Analysis, 28 (2009): 188-217.

[12] Novikov, A., *Optimal sequential tests for two simple hypothesis based on independent observations*, International Journal of Pure and Applied Mathematics, 25 (2008): 291-314.

[13] Novikov, A., *Una Introducción elemental al análisis estadístico secuencial*, Notas de curso. Primer coloquio del departamento de Matemáticas UAM-I, la Trinidad, Tlaxcala, México, 2007.

[14] Rockafeller, R. Tyrrell, *Convex Analysis*, Second Printing, Princeton University Press, 1972.

[15] Schmitz,N. *Optimal Sequentially Planned Decision Procedures*, Lecture Notes in Statistics 79, New York: Springer-Verlag, 1993.

[16] Shiryaev, A. N., *Optial Stopping Rules*, Springer-Verlag: Springer series in stadistics, New York, (1978), Reimpr. 2008.

[17] Siegmund, David, *Sequential Analysis: Tests and Confidence Intervals*, Springer-Verlag: Springer series in stadistics, New York, 1985.

[18] Wald, A., *Sequential Analysis*, Wiley, New York, 1947.

[19] Wald, A., *Sequential Test of Statistical Hypotheses*, Annals of Mathematical Statistics 16 (1945): 117-186.

[20] Wald, A., Wolfowitz, J., *Optimum character of the sequential probability ratio test*, Annals of Mathematical Statistics 19 (1948): 326-339.

[21] Xiong, X., Tan, M., y Boyett, J., *Sequential conditional probability ratio test for normalized test statistic on information time*, Biometrics 59 (2003): 624-631.