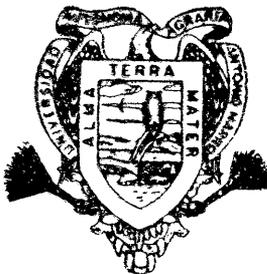


NORMALIDAD ASINTOTICA DE LA COTA MINIMAX
DEL OPTIMO DE UNA FUNCION DE
SUPERFICIES DE RESPUESTA

HECTOR FRANCO LOPEZ

T E S I S

PRESENTADA COMO REQUISITO PARCIAL
PARA OBTENER EL GRADO DE
MAESTRO EN CIENCIAS
EN ESTADISTICA EXPERIMENTAL



Universidad Autónoma Agraria
Antonio Narro

PROGRAMA DE GRADUADOS

Buenavista, Saltillo, Coah.

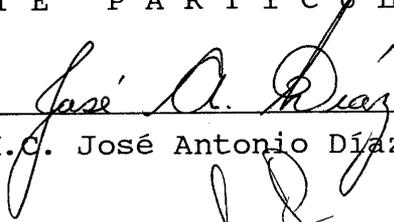
AGOSTO 1991

Tesis elaborada bajo la supervisión del comité particular de asesoría y aprobada como requisito parcial, para optar al grado de

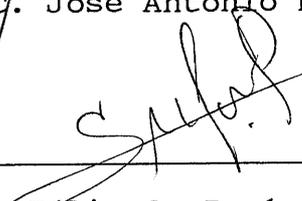
MAESTRIA EN CIENCIAS EN
ESTADISTICA EXPERIMENTAL

COMITE PARTICULAR

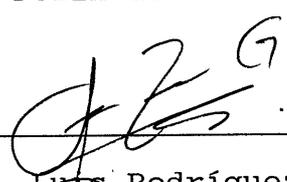
Asesor principal:

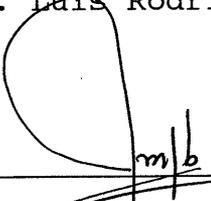

M.C. José Antonio Díaz García

Asesor:


M.C. Félix de Jesús Sánchez Pérez

Asesor:


M.C. Luis Rodríguez Gutiérrez


Dr. José Manuel Fernández Brondo

Subdirector de Asuntos de Postgrado



Buenavista, Saltillo, Coahuila. Agosto de 1991

AGRADECIMIENTOS

Agradezco a la Universidad Autónoma Agraria "Antonio Narro" todas las facilidades proporcionadas para poder realizar mis estudios de Posgrado.

Quiero hacer patente mi más profundo agradecimiento al Ing. M. en C. José Antonio Díaz García, Asesor de este trabajo de Tesis, por su empeñoso esfuerzo para la realización del mismo, sin el cual, su conclusión sería aún una lejana meta. Por su amistad gracias.

Agradezco a los Ingenieros M. en C. Félix de Jesús Sánchez P. y Luis Rodríguez G. por su apoyo en la revisión de este trabajo.

A mis compañeros de estudios, Gabriela Grijalba Monteverde y Leopoldo Ocegueda Altamirano, por su valiosa ayuda.

Por último no puedo dejar de reconocer la labor de todos mis maestros, en especial la del Dr. Rolando Cavazos C. A todos Gracias.

A MAGDALENA SOFIA

MAGDALENA Y

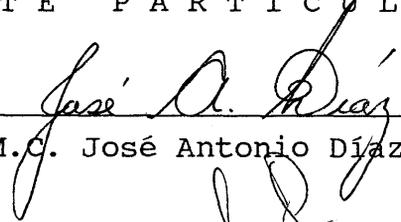
SOFIA

Tesis elaborada bajo la supervisión del comité particular de
asesoría y aprobada como requisito parcial, para optar al
grado de

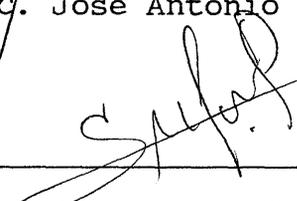
MAESTRIA EN CIENCIAS EN
ESTADISTICA EXPERIMENTAL

COMITE PARTICULAR

Asesor principal:

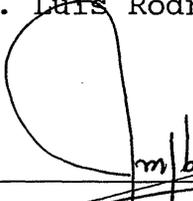

M.C. José Antonio Díaz García

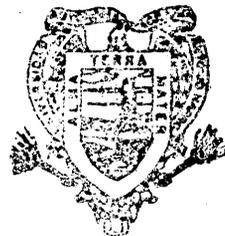
Asesor:


M.C. Félix de Jesús Sánchez Pérez

Asesor:


M.C. Luis Rodríguez Gutiérrez


Dr. José Manuel Fernández Brondo
Subdirector de Asuntos de Postgrado



BIBLIOTECA
EGIDIO G. REBONATC
BANCO DE TESIS
U.A.A.N.

Buenavista, Saltillo, Coahuila. Agosto de 1991

AGRADECIMIENTOS

Agradezco a la Universidad Autónoma Agraria "Antonio Narro" todas las facilidades proporcionadas para poder realizar mis estudios de Posgrado.

Quiero hacer patente mi más profundo agradecimiento al Ing. M. en C. José Antonio Díaz García, Asesor de este trabajo de Tesis, por su empeñoso esfuerzo para la realización del mismo, sin el cual, su conclusión sería aún una lejana meta. Por su amistad gracias.

Agradezco a los Ingenieros M. en C. Félix de Jesús Sánchez P. y Luis Rodríguez G. por su apoyo en la revisión de este trabajo.

A mis compañeros de estudios, Gabriela Grijalba Monteverde y Leopoldo Ocegueda Altamirano, por su valiosa ayuda.

Por último no puedo dejar de reconocer la labor de todos mis maestros, en especial la del Dr. Rolando Cavazos C. A todos Gracias.

A MAGDALENA SOFIA

MAGDALENA Y

SOFIA

COMPENDIO

Normalidad Asintótica de la Cota Minimax del
Óptimo de una Función de Superficie de respuesta

POR

HECTOR FRANCO LOPEZ

MAESTRIA

ESTADISTICA EXPERIMENTAL

UNIVERSIDAD AUTONOMA AGRARIA ANTONIO NARRO
BUENAVISTA, SALTILLO, COAHUILA. AGOSTO 1991.

M.C. José Antonio Díaz García

Palabras clave: Optimización estocástica, normalidad asintótica, técnica minimax, superficie de respuesta.

En el presente trabajo se replantea el problema de optimización de una función de superficie de respuesta como un programa de optimización estocástica, aplicando la técnica minimax desarrollada por Záčková. Así mismo se demuestra, aplicando técnicas convencionales, que el óptimo del programa estocástico es un estimador consistente que converge asintóticamente a la normal. La prueba de este último resultado se desarrolla apoyándose en la técnica de

minimización secuencial sin restricciones, SUMT, aprovechando la característica de convexidad -concavidad del problema.

ABSTRACT

Asymptotic Normality of the Optimum of a Response
Surface Function using Minimax approach

BY

HECTOR FRANCO LOPEZ

MAESTRIA

ESTADISTICA EXPERIMENTAL

UNIVERSIDAD AUTONOMA AGRARIA ANTONIO NARRO
BUENAVISTA, SALTILLO, COAHUILA. AGOSTO 1991.

M.C. José Antonio Díaz García

Key words: Stochastic programing, Asymptotic normality, Minimax technique, Respose surface.

This work state the problem of response surface optimization like stochastic programing, applying the minimax approach developed by Záčková. Thus was prove, doing it with conventional techniques, that the optimum of the stochastic programing problem is a consistent estimator that converges to a normal distribution. The proof of this last, was developed making use of sequential unconstrained minimization technique (SUMT), because the concavity -convexity of the problem.

INDICE DE CONTENIDO

	Página
1. INTRODUCCION	1
2. PRELIMINARES	5
2.1 SUPERFICIES DE RESPUESTA	5
2.2 OPTIMIZACION DE SUPERFICIES DE RESPUESTA.....	6
2.2.1 ANALISIS DE SUPERFICIE AJUSTADA	7
2.2.2 METODO DE ASCENSO POR LA MAXIMA PENDIENTE	10
2.2.3 ANALISIS DE CORDILLERA	11
2.3 PROGRAMACION CONVEXA	12
2.4 OPTIMIZACION MATEMATICA ESTOCASTICA..	17
2.4.1 TECNICA DE PROGRAMACION DE RESTRICCIONES PROBABI-- LISTICAS	18
2.4.2 TECNICA MINIMAX	20
2.5 LA DISTRIBUCION NORMAL MULTIDIMENSIONAL	24
2.5.1 FUNCION DE DENSIDAD	28
2.6 ALGUNOS TOPICOS DE CONVERGENCIA	28
3. PLANTEAMIENTO DEL PROBLEMA Y NORMALIDAD ASINTOTICA	36
3.2 METODOS PENALES	38

3.2.1 SUMT	39
3.3 NORMALIDAD ASINTOTICA	43
4. LITERATURA CITADA	54
APENDICE A	57
APENDICE B	61

1. INTRODUCCION

Gran cantidad de trabajos de investigación que se planean y llevan a cabo en la actualidad, tienen como objetivo primordial conocer el comportamiento o respuesta de una característica al aplicarse sobre ella un estímulo, caracterizado por un conjunto de variables controlables por el investigador. Una de las metodologías de investigación más comúnmente empleadas con este propósito, es la metodología de la Superficie de Respuesta, la cual consiste de una serie de técnicas que permiten la planeación y ejecución de una experimentación secuencial, donde se toman los resultados de la experimentación anterior en un proceso continuo. Conjetura - Diseño de Experimentos - Experimentacion - Análisis - Conjetura, y que permite establecer los niveles de estímulo en las variables controlables que optimicen la respuesta de la variable de interés.

La metodología de la superficie de respuesta consta de los siguientes aspectos:

- 1) Diseño de experimentos tendiente a la búsqueda de la respuesta óptima en la variable de interés.

- 2) Determinación del modelo que proporcione un mayor ajuste a los datos obtenidos en el punto anterior.

3) Encontrar el nivel, de las variables controlables, que optimice la respuesta de la variable de interés.

Los aspectos considerados en el primer inciso, han sido los más ampliamente abordados por los investigadores, dando como resultado grandes progresos en el desarrollo de la estrategia experimental de la metodología. Las técnicas de regresión lineal y no lineal resuelven lo planteado en el apartado número dos, encontrando el modelo estocástico que proporcione mayor ajuste a los datos de la experimentación. El presente estudio se ubica exclusivamente en el tercer inciso y tiene como objetivo general proponer una metodología de optimización que, a diferencia de las técnicas actuales, considere la aleatoriedad de los coeficientes de regresión del modelo. Hasta el momento las técnicas generalmente empleadas en la optimización de este tipo de problemas consideran determinísticos a los modelos encontrados.

A partir de lo anterior se plantean los siguientes objetivos particulares:

a) Replantear el problema de optimización considerando aleatorios los coeficientes de regresión del modelo.

b) Proponer una técnica que resuelva el problema anterior.

c) Establecer la convergencia en probabilidad y distribución de la solución obtenida bajo esta nueva técnica.

El presente estudio se encuentra dividido en tres capítulos principales. A continuación se describe brevemente el contenido temático de los capítulos subsecuentes.

El capítulo número dos plantea los preliminares del estudio, en el se aborda el problema de optimización en la metodología de la superficie de respuesta y analiza brevemente la manera actual de darle solución. A continuación se trata el punto de programación convexa, en el que principalmente se dan las bases para la caracterización de un punto óptimo dentro de una región de factibilidad convexa. El siguiente subcapítulo se refiere a la optimización matemática estocástica y en particular a la técnica minimax, escogida como enfoque en la solución del problema de optimización. Por último se estudian las características de la distribución normal y algunos tópicos de convergencia que serán auxiliares en la determinación de la distribución asintótica del óptimo.

En el capítulo número tres se replantea el problema de optimización como un problema estocástico, además de encontrar la distribución asintótica de la solución, empleando la técnica minimax auxiliada por la técnica de minimización secuencial sin restricciones (SUMT por sus siglas en Inglés).

Cabe hacer mención que la mayoría de los estudios respecto a este tópico emplean para su solución la técnica de los multiplicadores de Lagrange, sin embargo esta presenta el inconveniente de que la solución queda en función de los multiplicadores, además de obligar al

cumplimiento de las condiciones de Khun-Tucker, las cuales la hacen impráctica. Es por lo anterior que se optó por la utilización del método penal SUMT.

Finalmente se presentan dos apéndices, referentes a álgebra de matrices y regresión.

2. PRELIMINARES

2.1. Superficies de Respuesta

Se asume que el investigador está relacionado con un sistema, el cual considera alguna respuesta η la cual depende de variables de entrada $\xi_1, \xi_2, \dots, \xi_k$. También se asume que las ξ 's variables pueden ser controladas por el experimentador con un mínimo de error.

En general se tiene que:

$$\eta = f(\xi_1, \xi_2, \dots, \xi_k)$$

en donde la forma de la función (f) es desconocida y tal vez sumamente complicada. El éxito de la metodología de la superficie de respuesta depende de la aproximación de f por un polinomio de bajo orden en alguna región para las variables independientes. (Myers, 1971)

El experimentador comúnmente se interesa en: 1). Encontrar una aproximación de la función con el propósito de predecir futuras respuestas; y 2). Determinar cuales son los valores de las variables independientes que optimizan la respuesta, dentro de algún rango de operación.

Como ya se mencionó en el capítulo introductorio los aspectos considerados en el primer inciso, han sido ampliamente abordados por los investigadores, dando como resultado grandes progresos en el desarrollo de la estrategia experimental de la metodología de la superficie de respuesta. (Myers 1971 y Khuri y Cornell, 1987). El interés del presente estudio se centra en la segunda fase.

La fase 2)., también conocida como "fase de optimización" del problema, trata de encontrar los valores de x_1, x_2, \dots, x_k que maximicen (o minimicen) la respuesta, en general intenta encontrar condiciones óptimas de operación.

2.2. Optimización de Superficies de Respuesta.

Para fines del presente trabajo la optimización será tratada siempre como un problema de maximización. Rao (1979), menciona que sin pérdida de generalidad, la optimización puede ser tomada siempre como un problema de maximización, ya que el mínimo de una función $f(x)$ puede ser encontrado maximizando la función $-f(x)$.

Comúnmente existen dos diferentes situaciones en el problema de localización del óptimo: 1). El caso en que el punto crítico (óptimo) se encuentra dentro de la región experimental y , 2) El óptimo se encuentra remoto a la región experimental. A continuación se presentan los métodos clásicos, empleados en la resolución de estos problemas.

2.2.1. Análisis de la Superficie Ajustada.

En esta sección se considera la situación en la cual el experimentador tiene seleccionada una región que considera vecina al punto óptimo. (Myers, 1971)

Suponga que para k variables independientes, se tiene un modelo de respuesta de segundo orden del tipo

$$\eta = \beta_0 + \sum_{j=1}^k \beta_j x_j + \sum_{\substack{j=1 \\ j < m}}^k \sum_{m=1}^k \beta_{jm} x_j x_m + \sum_{j=1}^k \beta_{jj} x_j^2 \quad (2.1)$$

donde la x_{iu} son variables estandarizadas definidas de la siguiente forma:

$$x_{iu} = \frac{\xi_{iu} - \bar{\xi}_i}{\delta_i}$$

donde:

$$\delta_i = \left[\sum_{u=1}^N \frac{[\xi_{iu} - \bar{\xi}_i]^2}{N/C} \right]^{1/2}$$

por lo que es claro que:

$$\sum_{u=1}^N x_{iu} = 0 \quad y \quad \sum_{u=1}^N x_{iu}^2 = N/C$$

donde generalmente $C = 1$. (Vivanco, 1988).

Por medio de las técnicas de regresión, las cuales se discuten en el Apéndice B, se obtiene la ecuación

estimada:

$$\hat{Y} = b_0 + \sum_{j=1}^k b_j x_j + \sum_{\substack{j=1 \\ j < m}}^k \sum_{m=1}^k b_{jm} x_j x_m + \sum_{j=1}^k b_{jj} x_j^2 \quad (2.2)$$

la cual puede ser representada matricialmente como:

$$\hat{Y} = Z^T(x) \hat{\beta} \quad (2.3)$$

donde:

$$Z^T(x) = \left[1, x_1, \dots, x_k, x_1^2, \dots, x_k^2, x_1 x_2, \dots, x_1 x_k, \dots, x_{k-1} x_k \right]$$

$$\hat{\beta}^T = \left[b_0, b_1, \dots, b_k, b_{11}, \dots, b_{kk}, b_{12}, \dots, b_{1k}, \dots, b_{k-1k} \right]$$

o alternativamente, como:

$$\hat{Y} = b_0 + x^T b + x^T B x \quad (2.4)$$

donde:

b_0 es como antes y.

$$x^T = (x_1, x_2, \dots, x_k)$$

$$b^T = (b_1, b_2, \dots, b_k)$$

$$B = \begin{bmatrix} b_{11} & b_{12/2} & \dots & b_{1k/2} \\ & b_{22} & \dots & b_{2k/2} \\ & & \dots & b_{k-1k/2} \\ & & & b_{kk} \end{bmatrix}$$

el punto estacionario $x_0^T = (x_{1,0}, x_{2,0}, \dots, x_{k,0})$ está dado por la solución del sistema:

$$\frac{\partial \hat{Y}}{\partial x} = \frac{\partial}{\partial x} [b_0 + x^T b + x^T B x] = 0$$

$$b + 2Bx = 0$$

$$x_0 = (-B^{-1}b) / 2 \quad (2.5)$$

Este punto estacionario puede tratarse de un máximo, un mínimo o un punto de silla. (Myers, 1971; Khuri y Cornell, 1987). Con el fin de caracterizar tal punto, en el análisis clásico de superficies de respuesta se utiliza la siguiente metodología.

Análisis Canónico.

El objetivo de este análisis es determinar la naturaleza del punto estacionario (x_0) y del sistema de respuesta, por medio de una transformación de la ecuación (2.3), de manera que ésta sea más fácilmente interpretada.

En si, el análisis canónico se reduce a verificar las condiciones de suficiencia del punto estacionario, esto

es, analizar la matriz de segundas derivadas

$$H(x) = \frac{\partial^2 Y(x)}{\partial x \partial x^T}$$

así si, $H(x) \geq 0$ se tiene que el punto x_0 es un mínimo. Por el contrario si $H(x) < 0$ nos encontramos en un máximo, en el caso de que $H(x)$ sea indefinida el punto es un punto de silla. Lo anterior se apoya en lo descrito en las definiciones A.6 y A.7 del apéndice. (Myers, 1971; Khuri y Cornell, 1987 y Vivanco, 1988)

2.2.2. Método de Ascenso por la Máxima Pendiente.

Este método se emplea cuando no se tiene un conocimiento a priori sobre la localización del punto óptimo o cuando dentro de la región experimental no fueron encontrados puntos estacionarios.

El procedimiento de ascenso por la pendiente máxima es un método de experimentación secuencial por medio del cual el investigador procede a realizar experimentación a lo largo del camino de máximo incremento en respuesta. Los pasos a seguir en este procedimiento son los siguientes:

- 1) El experimentador ajusta un modelo de respuesta de primer orden en alguna región restringida para las variables x_1, x_2, \dots, x_k .

2) La información del paso 1 se emplea para localizar el camino de máxima pendiente.

3) Se lleva a cabo una serie de experimentos a lo largo del camino, hasta que el incremento en respuesta no sea evidente.

4) Se repiten los pasos 1, 2 y 3, usando la nueva región.

5) Si la curvatura es evidente y el experimentador considera que no puede obtenerse información adicional, se lleva a cabo un experimento más elaborado tendiente a la búsqueda de un modelo de segundo orden y se complementa con Análisis Canónico. (Vivanco, 1988 y Myers, 1971)

2.2.3. Análisis de Cordillera.

Durante el análisis de la superficie de respuesta ajustada, puede ocurrir que el punto estacionario no se encuentre dentro de la región experimental, y es deseable encontrar el valor de respuesta óptima dentro de la región. La búsqueda de ese valor óptimo de \hat{Y} es posible por medio del análisis de cordillera. El objetivo de esta técnica es el encontrar el valor máximo de \hat{Y} en la región experimental. (Khuri y Cornell, 1987)

Asumamos que el modelo ajustado sobre la región de las variables codificadas x_1, x_2, \dots, x_k , es de segundo orden y de la forma:

$$\hat{Y} = b_0 + \mathbf{x}^T \mathbf{b} + \mathbf{x}^T \mathbf{B} \mathbf{x}$$

donde b_0 , x y B son como antes. Suponga que se restringe la búsqueda del punto estacionario a una esfera de radio R , esto es, se restringe la búsqueda del óptimo a la condición:

$$\sum_{i=1}^k x_i^2 = R^2$$

Una vez que las coordenadas del óptimo son encontradas para un valor particular de R , se puede cambiar este valor y repetir el procedimiento. Al repetir el procedimiento escogiendo diferentes valores de R y graficando estos contra las coordenadas apropiadas de x_1 , x_2 , ..., x_k , y \hat{Y} , se producen gráficos del valor máximo de \hat{Y} para varias distancias del centro del diseño. (Myers, 1971; Khuri y Cornell, 1987). Un procedimiento alternativo es utilizar la técnica de "Nelder and Mead". (Khuri y Cornell, 1987).

2.3. Programación Convexa.

Definición 2.1. Programación Convexa. Un programa de optimización convexo es aquel que maximiza (minimiza) una función cóncava (convexa) dentro de una región de factibilidad convexa. (Prawda, 1982).

A continuación se verán algunas definiciones y resultados concernientes a la programación convexa.

Definición 2.2. Funciones Cóncavas y Convexas. Se dice que una función $f(X)$ es convexa en una región de X si ,

para cualquier par de puntos X_1, X_2 en la región y para toda $\lambda, 0 \leq \lambda \leq 1$,

$$f[\lambda X_2 + (1 - \lambda)X_1] \leq \lambda f(X_2) + (1-\lambda)f(X_1).$$

Similarmente, una función $f(X)$ se dice que es cóncava, en una región de X , si para cualquier par de puntos X_1, X_2 en la región y si para toda $\lambda, 0 \leq \lambda \leq 1$,

$$f[\lambda X_2 + (1 - \lambda)X_1] \geq \lambda f(X_2) + (1-\lambda)f(X_1).$$

A partir de estas definiciones es claro que si $f(X)$ es cóncava, entonces $-f(X)$ es convexa y viceversa. (Cooper y Steinberg, 1970).

De manera intuitiva, una función es convexa si una línea trazada entre cualquier par de puntos, en su gráfica, cae enteramente sobre o arriba de la gráfica. En el caso de las funciones cóncavas esta línea quedará sobre o por debajo de la gráfica. (Rao, 1979).

Teorema 2.1. Una función $f(X)$ es convexa si, para cualquier par de puntos X_1, X_2 se tiene

$$f(X_1) \geq f(X_2) + \nabla f^T(X_2)(X_1 - X_2) \quad (2.6)$$

Si $f(X)$ es cóncava, la desigualdad opuesta en (2.6) es válida. (Rao, 1979, p.687)

De la expansión en series de Taylor se tiene que

$$f(X_1) = f(X_2) + \nabla f^T(X_2)(X_1 - X_2) + \frac{1}{2}(X_1 - X_2)^T H(X_2 + \lambda(X_1 - X_2))(X_1 - X_2)$$

o sea

$$f(X_1) - f(X_2) - \nabla f^T(X_2)(X_1 - X_2) = \frac{1}{2}(X_1 - X_2)^T H(X_2 + \lambda(X_1 - X_2))(X_1 - X_2)$$

donde:

$$H(X) = \frac{\partial^2 f(X)}{\partial X \partial X'}$$

del Teorema 2.1 se sabe que

$$f(X_1) - f(X_2) - \nabla f^T(X_2)(X_1 - X_2) \geq 0$$

por lo que

$$0 \leq \frac{1}{2}(X_1 - X_2)^T H(X_2 + \lambda(X_1 - X_2))(X_1 - X_2)$$

para todo X_1 , X_2 y λ como antes. Por lo tanto $H \geq 0$ (Positiva semidefinida).

Así el siguiente teorema muestra la manera práctica de determinar la convexidad o concavidad de una función $f(X)$.

Teorema 2.2. Una función $f(X)$ es convexa (cóncava) si la matriz Hessiana

$$H(X) = \frac{\partial^2 f(X)}{\partial X \partial X'}$$

es positiva semidefinida (negativa semidefinida).

Teorema 2.3. Si la función $g(X)$ es convexa, entonces la región

$$R = \{X | g(X) \leq k\}$$

es convexa para todos los valores del parámetro k . Si $g(x)$ es cóncava, entonces la región

$$R = \{X | g(X) \geq k\}$$

es convexa para todos los valores del parámetro k .

Prueba. La prueba del teorema se hará para cuando la función $g(X)$ sea convexa, ya que cuando $g(X)$ es cóncava la prueba es análoga.

Dados dos puntos $X_1 = [X_{11}, X_{12}, \dots, X_{1n}]^T$ y $X_2 = [X_{21}, X_{22}, \dots, X_{2n}]^T$ en el conjunto R , se tiene por la convexidad de $g(x)$ que

$$g[\lambda X_2 + (1 - \lambda)X_1] \leq \lambda g(X_2) + (1-\lambda)g(X_1)$$

para $0 \leq \lambda \leq 1$. Pero por las restricciones impuestas en el conjunto R , se tiene que $g(X_1) \leq k$ y $g(X_2) \leq k$. Por lo tanto

$$g[\lambda X_2 + (1 - \lambda)X_1] \leq \lambda g(X_2) + (1-\lambda)g(X_1) \leq \lambda k + (1-\lambda)k = k$$

Esto implica que el punto $\lambda X_2 + (1 - \lambda)X_1$ también está en el conjunto R , para cualquier valor de λ . Por lo tanto R es un conjunto convexo.

Teorema 2.4. Cada punto mínimo (máximo) local de un programa de optimización convexo es un mínimo (máximo) global.

Prueba. Este teorema se probará por contradicción. Suponga que existen 2 mínimos locales, sean estos X_1 y X_2 , para la función $f(X)$. Sea $f(X_1) < f(X_2)$. Debido a que $f(X)$ es convexa la ecuación (2.6) se cumple y

$$f(X_1) - f(X_2) \geq \nabla f^T(X_2)(X_1 - X_2)$$

así

$$\nabla f^T(X_2)S \leq 0 \quad (2.7)$$

donde $S = (X_1 - X_2)$. La ecuación (2.7) indica que el valor de la función $f(x)$ decrece al moverse a partir del punto X en la dirección S . Esta conclusión es contradictoria con la asunción de que X_2 es un mínimo local. Por lo tanto no existe más que un solo mínimo para una función convexa. En el caso de una función cóncava la prueba es análoga.

El siguiente resultado juega un papel de suma importancia en el presente trabajo.

Teorema 2.5. Sea R una región convexa y sea $f(X)$ una función convexa sobre R en \mathbb{R}^n . Si $f(X)$ tiene un máximo

global en el punto X_0 , con $\|X_0\|$ finito, entonces uno o más X_0 se encuentran en la frontera de la región. (Cooper y Steinberg, 1970).

Un caso particular de este resultado es cuando f es un hiperplano, en este caso si la región factible es convexa, este teorema asegura que f obtiene su óptimo en la frontera de la región.

2.4. Optimización Matemática Estocástica

La programación estocástica aborda situaciones donde algunos o todos los parámetros, o inclusive las variables de decisión, de la optimización son variables aleatorias más que cantidades determinísticas. (Rao, 1979).

La programación estocástica es una rama de las matemáticas dado que trabaja con problemas de programación matemática, donde algunos de los parámetros son variables aleatorias. Al mismo tiempo debemos de considerarla como una parte o extensión de la estadística, dado que los problemas concernientes a la optimización de sistemas estocásticos en general pueden considerarse pertenecientes a la estadística, esto debido a que en un sentido amplio, algunas definiciones de estadística incluyen la posibilidad de obtener información por medio de la experimentación hacia un objetivo que depende de un estadístico. (Prékopa, 1978).

La idea básica al resolver un problema de optimización estocástica, es el convertir este en otro

problema equivalente determinístico. El problema resultante podrá ser resuelto usando técnicas, más familiares, como las de programación lineal o no lineal.

Existen diversas técnicas empleadas en la resolución de este tipo de problemas de optimización. Las más usuales han sido: La técnica de programación matemática bajo incertidumbre de Dantzig (Prawda, 1981), la técnica de programación de 2 etapas de Madansky (Rao, 1979), la técnica de programación de restricciones probabilísticas de Charnes-Cooper (Charnes y Cooper, 1959 y Rao 1979) y la técnica Minimax propuesta por Záčková (Dupačová, 1984).

Para los fines del presente trabajo solo serán descritas las técnicas Minimax y de Charnes-Cooper, la primera ya que será el enfoque utilizado en la solución de el problema de optimización y la segunda, ya que es la forma más conocida de resolver los problemas de programación estocástica. (Rao, 1979).

2.4.1. Técnica de Programación de Restricciones probabilísticas.

Esta técnica se emplea en la solución de problemas en los cuales las restricciones tienen cierta probabilidad de ser violadas. La técnica de programación de restricciones probabilísticas fué originalmente desarrollada por Charnes y Cooper.

En esta técnica el problema de programación lineal puede ser planteado como:

Minimice

$$f(X) = \sum_{j=1}^n c_j x_j \quad (2.8)$$

sujeto a:

$$P \left[\sum_{j=1}^n a_{ij} x_j \leq b_i \right] \geq p_i, \quad i = 1, 2, \dots, m \quad (2.9)$$

y

$$x_j \geq 0, \quad j = 1, 2, \dots, n$$

donde c_j , a_{ij} y b_i son variables aleatorias y p_i son probabilidades especificadas.

Note que las ecuaciones (2.9) indican que la i -ésima restricción,

$$\sum_{j=1}^n a_{ij} x_j \leq b_i$$

debe ser satisfecha con una probabilidad de al menos p_i donde $0 \leq p_i \leq 1$. Por simplicidad, se asume que las variables de decisión x_j son determinísticas.

Programación Estocástica no Lineal.

El problema de programación no lineal estocástica puede ser planteado como:

$$\text{Encuentre } X \text{ tal que minimize } f(Y) \quad (2.10)$$

tal que

$$P[g_j(Y) \geq 0] \geq p_j, \quad j = 1, 2, \dots, m \quad (2.11)$$

donde Y es el vector de N variables aleatorias Y_1, Y_2, \dots, Y_n y este incluye las variables de decisión x_1, x_2, \dots, x_n . Para fines del presente se asume que todas las variables aleatorias son independientes y distribuidas normalmente. Las ecuaciones (2.11) denotan que la probabilidad de que $g_j(Y)$ sea mayor o igual a cero debe de ser mayor o igual que la probabilidad p_j especificada. El problema planteado en las ecuaciones (2.10) y (2.11) puede ser convertido en uno equivalente determinístico aplicando la técnica de programación de restricciones probabilísticas. (Rao, 1979).

2.4.2. Técnica Minimax.

Considere el siguiente problema de optimización estocástica con restricciones

$$\text{maximize } E_{\theta} \left\{ c^T x - \varphi(x; A, b) \right\} \text{ en el conjunto } \mathcal{H} \quad (2.12)$$

donde \mathcal{H} es el conjunto de soluciones admisibles. Un ejemplo de (2.12) es cuando el programa lineal

$$\text{maximize } c^T x$$

$$\text{sujeto a } Ax \leq b, x \geq 0,$$

tiene algunos elementos del vector $b \in \mathbb{R}^m$, vector $c \in \mathbb{R}^n$ o de $A \in \mathbb{R}^{m \times n}$ aleatorios.

Asuma

(i) Para A y b fijas, $\varphi(x; A, b)$ es una función convexa no negativa de x .

(ii) Para una x arbitraria tal que $x \in \mathcal{H}$, $\varphi(x; A, b)$ es una función convexa de A y b .

(iii) $\mathcal{H} \subset \mathbb{R}^n$ es un conjunto convexo cerrado y no vacío.

Si se supone que la distribución conjunta de los coeficientes aleatorios θ es conocida, entonces el problema planteado en (2.12) es en principio reducible a un problema determinístico no lineal. Este tipo de problemas han sido estudiados por diferentes autores y bajo distintos puntos de vista. (Armitano, Edelman y García, 1985; Cooper y Steinberg, 1970; Prawda, 1982 y Rao, 1979). Su forma explícita así como su solución óptima depende de la distribución de θ . Ante esto deben de considerarse dos alternativas:

I. La distribución de θ pertenece a una familia de distribuciones paramétricas dada.

II. La distribución de θ pertenece a un conjunto de distribuciones especificado definido por valores prescritos de algunos momentos.

El problema que se encara en el caso I es el de resolver el problema

$$\text{maximice } f(x;\eta) \text{ en el conjunto } \mathcal{H} \quad (2.13)$$

donde:

$$f(x;\eta) = E_{\theta_{\eta}} \left\{ c^T x - \varphi(x;A,b) \right\} \quad (2.14)$$

y η es el vector de parámetros verdaderos de la distribución de θ . si η no se conoce precisamente, se substituye por un estimador, digamos ψ , y el programa (2.13) se substituye por el siguiente

$$\text{maximice } f(x,\psi) \text{ en el conjunto } \mathcal{H} \quad (2.15)$$

En el caso II, se admite que el conocimiento de la distribución θ es incompleto, pero limitado al hecho de que la distribución de θ pertenece a un conjunto de distribuciones Θ dado. Un acercamiento es vía Minimax; Cualquier solución óptima de

$$\text{maximice } f(x) = \min_{\theta \in \Theta} E_{\theta} \left\{ c^T x - \varphi(x;A,b) \right\} \quad (2.16)$$

deberá ser llamada como la solución minimax al problema de

optimización estocástica (2.12)

Como un resultado de esta técnica se obtienen las siguientes desigualdades, las cuales son relativamente fáciles de calcular:

$$\begin{aligned} \max_{x \in \mathcal{H}} \min_{\theta \in \Theta_\eta} E_\theta \left\{ c^T x - \varphi(x; A, b) \right\} &\leq \max_{x \in \mathcal{H}} E_\theta \left\{ c^T x - \varphi(x; A, b) \right\} \\ &\leq \max_{x \in \mathcal{H}} \max_{\theta \in \Theta_\eta} E_\theta \left\{ c^T x - \varphi(x; A, b) \right\} \quad \forall \theta \in \Theta_\eta \end{aligned} \quad (2.17)$$

La forma explícita de la función objetivo

$$f(x; \eta) = \min_{\theta \in \Theta_\eta} E_\theta \left\{ c^T x - \varphi(x; A, b) \right\}$$

y la solución óptima de (2.16) depende del vector de parámetros η . Cuando los valores prescritos η de los momentos no se conocen de manera precisa, lo cual comunmente es la situación, el valor de η puede ser sustituido por ψ y, de manera similar al caso I, resolver el siguiente problema sustituto

$$\max_{x \in \mathcal{H}} f(x; \psi) = \min_{\theta \in \Theta_\psi} E_\theta \left\{ c^T x - \varphi(x; A, b) \right\} \quad (2.18)$$

en lugar de $\max_{x \in \mathcal{H}} f(x; \eta)$

La parte central de la desigualdad (2.17) muestra la forma clásica de resolver los problemas de optimización, más

aún, en los problemas de optimización de superficies de respuesta ni siquiera se considera el optimizar la esperanza de la función de respuesta. La primera parte de la ecuación (2.18) muestra una cota mínima para los máximos (el mínimo de los máximos) considerando la aleatoriedad del problema, la cual es una fuerte herramienta en la toma de decisiones, ya que de alguna manera incrementa la confianza en el valor óptimo.

Motivados en la discusión anterior, dicha cota es utilizada como fundamento de la técnica y caracterización de la solución al problema de la función ESTOCASTICA de superficies de respuesta

2.5. La Distribución Normal Multidimensional.

Definición 2.3. Un vector aleatorio p -dimensional Y se dice que tiene distribución normal multidimensional $N_p(\mu, \Sigma)$ si Y tiene la misma distribución que $\mu + Bu$, con $B \in \mathbb{R}^{p \times p}$, donde el rango de B , $r(B) = p$, $\Sigma = BB^T$, y $Y \sim N_p(\mu, I_p)$. Siendo p es el rango de la distribución.

Teorema 2.6. Sea $Y \sim N_p(\mu, \Sigma)$, entonces:

i). $E(Y)$ y $\text{Var}(Y)$ existen y se denotan respectivamente por μ y Σ . (Rao, 1973).

ii). La función característica de Y es

$$\phi_Y(t) = \exp\left[it^T\mu - \frac{1}{2} t^T\Sigma t\right] \quad (2.19)$$

(Srivastava y Khatri, 1979)

Teorema 2.7. Sea $a \in \mathbb{R}^p$ de constantes, $a \neq 0$, entonces $a^T Y \sim N(a^T \mu, a^T \Sigma a)$ si y sólo si $Y \sim N_p(\mu, \Sigma)$.

Esto es cierto aún si $a^T \Sigma a = 0$, en cuyo caso la distribución es conocida como degenerada, dado que toda la masa se concentra en $a^T \mu$.

Prueba. Supongamos que $Y \sim N_p(\mu, \Sigma)$ y veamos que $a^T Y \sim N(a^T \mu, a^T \Sigma a)$

Considere el caso más general. Sea $X = A^T Y$ donde $A^T \in \mathbb{R}^{s \times p}$, $s < p$ y $r(A) = s$. Entonces

$$\phi_x(t) = E\left[\exp(it^T x)\right]$$

como $x = A^T Y$

$$\phi_{A^T Y}(t) = E\left[\exp(it^T A^T Y)\right] = E\left[\exp(i[At]^T Y)\right] = \phi_Y(At)$$

como $Y \sim N_p(\mu, \Sigma)$, del teorema 2.6. ii., se tiene:

$$\begin{aligned} \phi_x(t) &= \phi_Y(At) = \exp\left[i(At)^T \mu - \frac{1}{2} (At)^T \Sigma (At)\right] \\ &= \exp\left[it^T (A^T \mu) - \frac{1}{2} t^T (A^T \Sigma A) t\right] \end{aligned}$$

de donde $x \sim N(A^T \mu, A^T \Sigma A)$.

Ahora recuerde que si $\phi_x(t)$, es la función característica de x con $x, t \in \mathbb{R}^p$, entonces $\forall i = 1, 2, \dots, p$ la función característica marginal de x_i está dada por la siguiente expresión:

$$\phi_{x_i}(t_i) = \phi_x\left[0, \dots, 0, t_i, 0, \dots, 0\right].$$

Así note que

$$A^T Y = \begin{bmatrix} a_1^T Y \\ \vdots \\ a_p^T Y \end{bmatrix}$$

con

$$A^T = \begin{bmatrix} a_1^T \\ \vdots \\ a_p^T \end{bmatrix}.$$

Luego $x_i = a_i^T Y$

$$\phi_{x_i}(t_i) = \exp \left[i t_i (a_i^T \mu) - \frac{1}{2} t_i^2 (a_i^T \Sigma a_i) \right] \quad \forall i = 1, 2, \dots, p$$

luego $a_i^T Y \sim N(a_i^T \mu, a_i^T \Sigma a_i) \quad \forall i = 1, 2, \dots, n$, lo cual prueba esta parte del teorema dado que a y A son arbitrarias.

Ahora supongamos que $a^T Y \sim N(a^T \mu, a^T \Sigma a)$ y veamos que $Y \sim N_p(\mu, \Sigma)$. Tome para todo $a \neq 0$

$$x_a = a^T Y$$

y observe que

$$E(x_a) = E(a^T Y) = a^T \mu \quad \text{y} \quad \text{Var}(x_a) = \text{Var}(a^T Y) = a^T \Sigma a$$

de esto, y por hipótesis para cada a , $x_a \sim N(a^T \mu, a^T \Sigma a)$, por lo tanto

$$\phi_{X_a}(t) = E\left[\exp(itx_a)\right] = \exp\left[it(a^T\mu) - \frac{1}{2}t^T(a^T\Sigma a)\right]$$

pero

$$E\left[\exp(itx_a)\right] = E\left[\exp(ita^TY)\right] = \phi_Y(ta)$$

como t y a son arbitrarios, se tiene que para $t = ta$

$$\phi_Y(t) = E\left[\exp(it^TY)\right] = \exp\left[t^T\mu - \frac{1}{2}t^T\Sigma t\right]$$

A partir del teorema anterior surge la siguiente definición alternativa de normalidad.

Definición 2.4. Un vector aleatorio p -dimensional Y se dice que tiene distribución normal multidimensional si toda combinación lineal no trivial de Y tiene distribución normal univariada (incluyendo la distribución normal degenerada). (Srivastava y Khatri, 1979).

El siguiente resultado asegura que toda forma cuadrática de vectores aleatorios normales tiene una distribución χ^2 .

Teorema 2.8. Sea $Y \sim N_p(0, I)$, entonces $Y^TY \sim \chi_p^2$.

Prueba. Solo observe que

$$Y^TY = \sum_{i=1}^p Y_i^2$$

con $Y_i \sim N(0, 1)$.

2.5.1. Función de Densidad.

La forma de función de Densidad de un vector aleatorio depende del rango de la matriz de varianzas y covarianzas Σ , $R(\Sigma)$. Así se presentan 2 situaciones distintas si la matriz de varianzas y covarianzas (Σ) es o no de rango completo. (Rao, 1973)

CASO 1. $R(\Sigma) = p$.

Sea $Y \sim N_p(\mu, \Sigma)$. Entonces

$$f_Y(Y) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp\left[-\frac{1}{2} (Y - \mu)^T \Sigma^{-1} (Y - \mu)\right],$$

la cuál es la densidad de Y .

Caso 2. $R(\Sigma) = k < p$.

Sea $Y \sim N_p(\mu, \Sigma)$. Entonces la densidad de Y puede ser escrita como:

$$f_Y(Y) = (2\pi)^{-k/2} [\lambda_1, \lambda_2, \dots, \lambda_k]^{-1/2} \exp\left[-\frac{1}{2}(Y-\mu)^T \Sigma^- (Y-\mu)\right]$$

donde Σ^- es el inverso generalizado de Σ y λ_i $i = 1, 2, \dots, k$ son las raíces características no nulas de la matriz Σ . Cabe destacar que esta representación no es única.

2.6. Algunos tópicos de Convergencia.

En esta sección se presentan algunas definiciones y resultados referentes al comportamiento asintótico de estimadores, tanto en probabilidad como en distribución.

Definición 2.5. (Consistencia). Sea $\{\hat{\theta}_n\}$, $n = 1, 2, \dots$, o simplemente $\hat{\theta}_n$ una sucesión de estimadores del parámetro vectorial θ . Se dice que $\hat{\theta}_n$ es una estimación consistente de θ si, para un δ dado y un $\xi > 0$, existe N tal que para $n > N$

$$P\left\{\|\hat{\theta}_n - \theta\| > \delta\right\} < \xi$$

o alternativamente

$$\lim_{n \rightarrow \infty} P\left\{\|\hat{\theta}_n - \theta\| > \delta\right\} = 0$$

donde $\|\hat{\theta}_n - \theta\|$ es la norma euclidiana en \mathbb{R}^n . Siendo así se dice que $\hat{\theta}_n$ converge en probabilidad a θ lo cual se representa también como $p\text{Lim } \hat{\theta}_n = \theta$ ó $\hat{\theta}_n \xrightarrow{p} \theta$.

Cabe hacer la aclaración de que el concepto de consistencia de la definición 2.5. es aplicable a una sucesión de estimadores y no a un estimador. Sin embargo es de uso común el hablar de estimador consistente lo cual será válido en el presente estudio con la salvedad anterior.

La primera situación a resolver es, dado un estimador, como saber si este es consistente o no. El siguiente resultado nos ayuda a responder a lo anterior. (Rao, 1973 y Cramer, 1970.).

Teorema 2.9. (Desigualdad de Chebyshev). Sea X una variable aleatoria de media μ y varianza σ^2 y sea $k > 0$. Entonces:

$$P\left\{|x - \mu| > k\right\} \leq \frac{\sigma^2}{k}$$

Para concluir con el estudio de consistencia se presenta el siguiente resultado sobre funciones de estimaciones consistentes, el cuál se probará siguiendo un razonamiento heurístico.

Teorema 2.10. Sea $\{X_n\}$ $n = 1, 2, \dots$ una sucesión de vectores aleatorios (estadísticos) que converge en probabilidad a la constante X (parámetro). Sea $g(\circ)$ una función continua tal que $g(X_n)$ está bien definida. Entonces la sucesión $\{g(X_n)\}$ $n = 1, 2, \dots$ converge en probabilidad a $g(X)$, siempre que $g(X)$ exista, esto es

$$\lim_{n \rightarrow \infty} P\left\{|g(X_n) - g(X)| > \delta\right\} = 0 \quad \text{ó}$$

$$P\left\{|g(X_n) - g(X)| > \delta\right\} \leq \xi \quad \forall n > N \quad \text{ó}$$

$$g(X_n) \xrightarrow{P} g(X)$$

Prueba. Sin pérdida de generalidad suponga que $X \in \mathbb{R}^k$. Hay que ver que dados ξ y $\delta > 0$, existe N tal que para todo $n > N$

$$P\left\{|g(X_n) - g(X)| < \delta\right\} \geq 1 - \xi$$

como $g(\circ)$ es continua, tome $\delta_1 > 0$ tal que si $|X_n - X| < \delta_1$

entonces $|g(X_n) - g(X)| < \delta$. Ahora como $X_n \xrightarrow{P} X$, dados δ_1 y $\xi > 0$, existe N tal que $\forall n > N$

$$P\left\{|X_n - X| < \delta_1\right\} > 1 - \xi$$

Así para todo X_n que verifique esta desigualdad, por la continuidad de $g(\circ)$ se tiene

$$|g(X_n) - g(X)| < \delta$$

Por lo tanto, $\forall n > N$

$$P\left\{|g(X_n) - g(X)| < \delta\right\} > P\left\{|X_n - X| < \delta_1\right\} \geq 1 - \xi$$

La distribución límite o distribución asintótica de una variable o vector aleatorio juega un papel muy importante dentro de la estadística matemática. Este caso se presenta cuando se calcula un estadístico $\hat{\theta}_n$ a partir de una muestra de tamaño n . En general es muy difícil encontrar la distribución exacta para todo n , en este caso la distribución real se aproxima por la distribución límite para valores grandes de n . Así formalmente se tiene:

Definición 2.6. (Convergencia en distribución). Se dice que la sucesión $\{X_n\}$ de variables (vectores) aleatorias converge en distribución a la variable X con función de distribución F si $\lim_{n \rightarrow \infty} F(X_n) = F(X)$ o simplemente $F_n \rightarrow F$ en todo punto de continuidad de F . Se denotará este hecho

escribiendo

$$X_n \xrightarrow{d} X$$

(Rao, 1973).

Dentro de las aplicaciones de la estadística el siguiente teorema de convergencia en distribución de una suma de variables (vectores) aleatorias, juega un papel preponderante en la inferencia de grandes muestras.

Teorema 2.11. (Teorema de Límite Central Multivariado). Sea la sucesión $\{X_n\}$ $n = 1, 2, \dots$ de vectores p -dimensionales aleatorios independientes e idénticamente distribuidos con $E(X_n) = \mu$ y $\text{Var}(X_n) = \Sigma \forall n = 1, 2, \dots$, entonces

$$Y = k^{-1/2} \sum_{n=1}^k (X_n - \mu) \xrightarrow{d} W$$

donde $W \sim N_p(0, \Sigma)$ con Σ como antes.

Prueba. Este teorema se puede probar esencialmente bajo los mismos lineamientos del caso univariado (Cramer, 1970), solo considere la combinación lineal $1^T X$ y la definición 2.4., donde $1^T = (1, 1, \dots, 1)^T \in \mathbb{R}^p$.

Otro problema común en la inferencia estadística es el determinar la distribución de la función, digamos $g(\circ)$, de un estadístico consistente, al igual que en el ejemplo expuesto anteriormente el encontrar la distribución exacta

de tal función es una tarea difícil. Para resolver este problema se considera (lo cual es así) que el estadístico $\hat{\theta}_n$ es consistente y función de una muestra $\{X_1, X_2, \dots, X_n\}$ y se procede a determinar la función de g en el límite. Cuando la distribución asintótica de $\hat{\theta}_n \xrightarrow{d} X \sim N(\mu, \Sigma)$. Los siguientes resultados nos ayudan a encontrar tal distribución asintótica. (Rao, 1973 y Cramer, 1970).

Teorema 2.12. Sea $(\hat{\theta}_n - \theta) \xrightarrow{d} Y \sim N(0, \Sigma)$, $\hat{\theta}_n \in \mathbb{R}^k$. Sea además $g: \mathbb{R}^k \rightarrow \mathbb{R}$ totalmente diferenciable (al menos 2 veces diferenciable) en un entorno del punto θ . Entonces la distribución asintótica de

$$\sqrt{n} \left[g(\hat{\theta}_n) - g(\theta) \right]$$

es una Normal con media cero y varianza

$$\left[\frac{\partial g(\theta)}{\partial \hat{\theta}} \right]^T \Sigma \left[\frac{\partial g(\theta)}{\partial \hat{\theta}} \right] \quad (2.20)$$

El lineamiento de la prueba es el siguiente. Desarrollando en series de Taylor se tiene

$$g(\hat{\theta}_n) = g(\theta) + \frac{\partial g(\theta)}{\partial \hat{\theta}} \left[\hat{\theta}_n - \theta \right] + R$$

donde R contiene derivadas de segundo orden. El primer término del segundo miembro es una constante, mientras que el segundo es la suma de n variables aleatorias independientes, cada una de las cuales tiene media cero. Por

los teoremas 2.7 y 2.11, podemos decir que, en condiciones generales, la suma de los dos primeros términos es asintóticamente normal, con media igual al primer término y matriz de varianzas y covarianzas dada por la ecuación (2.20).

Veamos ahora una generalización del teorema anterior.

Teorema 2.13. Sea $(\hat{\theta}_n - \theta) \xrightarrow{d} Y \sim N(0, \Sigma)$, $\hat{\theta}_n \in \mathbb{R}^k$. Sea además $g: \mathbb{R}^k \rightarrow \mathbb{R}^r$ totalmente diferenciable (al menos 2 veces) en un entorno del punto θ . Entonces la distribución asintótica de

$$\sqrt{n} \left[g(\hat{\theta}_n) - g(\theta) \right] \quad (2.21)$$

es una normal con media cero y varianza

$$G \Sigma G^T \quad (2.22)$$

donde

$$G = \begin{bmatrix} \frac{\partial g_1}{\partial \hat{\theta}_1} & \frac{\partial g_1}{\partial \hat{\theta}_2} & \dots & \frac{\partial g_1}{\partial \hat{\theta}_k} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \frac{\partial g_r}{\partial \hat{\theta}_1} & \frac{\partial g_r}{\partial \hat{\theta}_2} & \dots & \frac{\partial g_r}{\partial \hat{\theta}_k} \end{bmatrix} \quad (2.23)$$

Prueba. Para probar este resultado unicamente se necesita considerar la función lineal $g = b_1 g_1 + \dots + b_r g_r$ y

aplicar el teorema 2.12, entonces la distribución asintótica de

$$\sqrt{n} \mathbf{b}^T \left[\mathbf{g}(\hat{\theta}_n) - \mathbf{g}(\theta) \right]$$

es normal con media cero y varianza

$$\mathbf{b}^T \mathbf{G} \Sigma \mathbf{G}^T \mathbf{b}$$

donde $\mathbf{b}^T = (b_1, \dots, b_r)$. Por la definición 2.4 se obtiene el resultado deseado.

El teorema anterior no es aplicable en la práctica, esto debido a que la matriz de varianzas y covarianzas $\mathbf{G} \Sigma \mathbf{G}^T$ depende de parámetros comúnmente desconocidos. Un resultado más potente se enuncia a continuación (Rao, 1973)

Teorema 2.14. Sean \mathbf{G}_n y Σ_n los valores de \mathbf{G} y Σ cuando $\theta = \hat{\theta}_n$. Además, sea F_n la densidad de $\sqrt{n} [\mathbf{g}(\hat{\theta}_n) - \mathbf{g}(\theta)]$ y H_n la densidad de una distribución normal q -variada con media cero y matriz de varianzas y covarianzas $\mathbf{G}_n \Sigma_n \mathbf{G}_n^T$. Entonces si Σ y \mathbf{G} son continuas en θ .

$$\lim_{n \rightarrow \infty} \sup_{x_1, \dots, x_q} (F_n - H_n) = 0$$

El resultado anterior muestra que F_n puede ser aproximada por una distribución normal q -variada con matriz de varianzas y covarianzas $\mathbf{G}_n \Sigma_n \mathbf{G}_n^T$, la cual involucra únicamente estimadores.

3. PLANTEAMIENTO DEL PROBLEMA Y NORMALIDAD

ASINTOTICA DEL OPTIMO

3.1. Planteamiento del problema.

Hasta el momento las técnicas generalmente empleadas en la optimización de este tipo de problemas consideran a los modelos encontrados, a partir de las técnicas de regresión, como determinísticos. Sin embargo con un sencillo análisis es fácil ver que el modelo empleado es aleatorio ya que depende de estimadores ($\hat{\beta}$ y $\hat{\sigma}^2$). Así basándonos en la sección 2.4. el problema

$$\text{Maximizar } Z^T(x)\hat{\beta}$$

$$\text{sujeto a } x^T x \leq C^2$$

se puede replantear como un problema de optimización estocástica de la siguiente manera:

$$\text{Maximizar } E(Z^T(x)\hat{\beta})$$

$$\text{sujeto a } x^T x \leq C^2$$

La forma escogida para atacar este problema está fundamentada en la sección 2.4.2. y es conocida como la técnica minimax. De esta forma el problema puede ser planteado alternativamente como:

$$\text{Max}_x \text{Min}_\theta E(Z^T(x)\hat{\beta})$$

Este problema puede ser separado en dos partes:

a). Minimizar $Z^T(x)\hat{\beta}$
 θ

donde:

$$\theta^T = [\beta^T, \hat{\sigma}^2]$$

sujeto a:

$$(\hat{\beta} - \beta)^T X^T X (\hat{\beta} - \beta) - p \hat{\sigma}^2 F \leq 0$$

$$\frac{(n-p) \hat{\sigma}^2}{\chi_1} - \sigma^2 \leq 0$$

$$\sigma^2 - \frac{(n-p) \hat{\sigma}^2}{\chi_2} \leq 0$$

$p = 1 + 2k + k(k-1)/2$, $k =$ número de variables independientes x_i s.

Además las restricciones especifican una región de confianza simultánea del parámetro θ . (Ver apéndice B)

b). Maximizar $Z^T(x)\hat{\beta}_0$
 x

sujeto a:

$$\mathbf{x}^T \mathbf{x} - C^2 \leq 0$$

donde $\hat{\beta}_0$ es la solución al inciso a).

Para resolver las optimizaciones anteriores, la mayoría de los autores en el tema (Dupacová, 1984 y Melaku, 1986) emplean la técnica de los multiplicadores de Lagrange. En este trabajo se propone el uso de una técnica alternativa a esta, dado que las soluciones empleando multiplicadores de Lagrange quedan en función de los multiplicadores (λ_i 's), además de que obliga a cumplir con las condiciones de Khun-Tucker (Cooper y Steinberg, 1970 y Armitano, Edelman y García, 1985) las cuales son totalmente imprácticas para resolver problemas de optimización. (Prawda, 1982).

Sin embargo la importancia de las condiciones de Khun-Tucker, radica en que sientan las bases para el desarrollo de métodos prácticos de optimización, mucho más eficientes. Uno de estos se detalla a continuación.

3.2. Métodos penales.

Los métodos penales restringidos tienen como principio básico transformar a un problema de optimización restringido en uno no restringido, incorporando de cierta forma las restricciones en la función objetivo. Se les llama métodos penales, por que utilizan un término con el que penalizan a la función objetivo. (Prawda, 1982).

Sea el problema de optimización de la forma

$$\left. \begin{array}{l} \text{minimice } f(X) \text{ sujeto a:} \\ g_j(X) \leq 0, j = 1, 2, \dots, m \end{array} \right\} \quad (3.1)$$

este puede ser transformado en un problema de minimización no restringida construyendo una función de la forma

$$\phi_k = \phi(X, r_k) = f(X) + r_k \sum_{j=1}^m G_j[g_j(X)] \quad (3.2)$$

donde G_j es alguna funcional de las restricciones g_j y r_k es una constante positiva conocida como el parámetro de penalización. El segundo término en el lado derecho de la ecuación (3.2) es llamado término de penalización. Si la minimización sin restricciones de la función ϕ se repite para una secuencia de valores del parámetro de penalización r_k , el cual se conoce en cada iteración, la solución converge a la del problema (3.1).

Por cierto, el método clásico de optimización de Lagrange es un método penal paramétrico. De todos los métodos penales la técnica de minimización secuencial sin restricciones (SUMT por sus siglas en Inglés) ha demostrado ser la más eficiente y práctica, a continuación se procede a exponerla.

3.2.1. SUMT.

El SUMT es un método penal paramétrico de punto interior, debido a que la sucesión de puntos

factibles X_0, \dots, X_s encontrados en cada iteración del método deben estar en la región factible (cumplir con las restricciones). Las funciones G_j generalmente usadas son de la forma:

$$G_j = -\frac{1}{g_j(X)} \quad (3.3)$$

y

$$G_j = \log[-g_j(X)]. \quad (3.4)$$

Para los propósitos del presente estudio se optó por utilizar la funcional (3.3) así, sustituyendo esta en la ecuación (3.2) se tiene que la función penal es de la forma

$$\phi(X, r_k) = f(X) - r_k \sum_{j=1}^m \frac{1}{g_j(X)} \quad (3.5)$$

(Rao, 1979)

A continuación se enuncia el resultado que nos asegura la convergencia al óptimo empleando el SUMT.

Teorema 3.1. Si la función

$$\phi(X, r_k) = f(X) - r_k \sum_{j=1}^m \frac{1}{g_j(X)}$$

se minimiza para una secuencia decreciente de valores de r_k , el mínimo no restringido de $\phi(X, r_k)$, X_k^* , converge a la solución óptima del problema restringido planteado en (3.1)

si $r_k \rightarrow 0$. (Rao, 1979 y Fiacco y McCormick, 1966).

Hasta el momento se ha visto que la minimización secuencial de

$$\phi(X, r_k) = f(X) - r_k \sum_{j=1}^m \frac{1}{g_j(X)} > 0$$

para una secuencia decreciente de valores r_k proporciona el mínimo X_k^* . A medida que $k \rightarrow \infty$, esos puntos X_k^* convergen al mínimo del problema propuesto en (3.1). En orden de asegurar la existencia de un mínimo global de $\phi(X, r_k)$ para todo valor positivo de r_k , ϕ debe ser una función de X estrictamente convexa. Si ϕ es convexa, entonces para toda $r_k > 0$, un mínimo único de $\phi(X, r_k)$ está dado por

$$\left[\partial \phi(X, r_k) / \partial X \right] = 0$$

Teorema 3.2. Si las funciones $f(X)$ y $g_j(X)$ son convexas y al menos una de las dos es estrictamente convexa, entonces la función $\phi(X, r_k)$ (3.5) será una función de X estrictamente convexa. (Rao, 1979 y Fiacco y McCormick, 1966).

Resultados análogos a los de los teoremas 3.1 y 3.2 pueden ser establecidos para el caso de maximización solo considere:

i) En el caso del teorema 3.1 tome la función penal a maximizar como

$$\phi(X, r_k) = f(X) + r_k \sum_{j=1}^m \frac{1}{g_j(X)}$$

el cambio se desprende de la siguiente observación:

ii) En el caso del teorema 3.2 considere $f(X)$ concava y como

$$- r_k \sum_{j=1}^m \frac{1}{g_j(X)}$$

es convexa

$$r_k \sum_{j=1}^m \frac{1}{g_j(X)}$$

es concava.

Ahora los dos problemas propuestos en la sección 3.1. pueden ser replanteados bajo esta técnica como sigue:

$$\text{a) Min}_{\theta} \phi(\hat{\theta}, r; \theta) = Z^T(x)\hat{\beta} - r \left[\frac{1}{g_1} + \frac{1}{g_2} + \frac{1}{g_3} \right] \quad (3.6)$$

con

$$g_1 = (\hat{\beta} - \beta)^T X^T X (\hat{\beta} - \beta) - p \hat{\sigma}^2 F \leq 0$$

$$g_2 = \frac{(n-p) \hat{\sigma}^2}{\chi_1} - \sigma^2 \leq 0$$

$$g_3 = \sigma^2 - \frac{(n-p) \hat{\sigma}^2}{\chi_2} \leq 0$$

$$b) \underset{x}{\text{Max}} \phi(x, r; \theta_0) = Z^T(x) \beta_0 + r \left[\frac{1}{g} \right] \quad (3.7)$$

donde

$$g = x^T x - c^2 \leq 0$$

Además observe que: para el caso de la minimización ϕ es una función convexa y en el inciso b) ϕ es una función cóncava, esto es $B \leq 0$, por el teorema 2.4 se obtiene el resultado deseado, si en su defecto $B \geq 0$ el teorema 2.5 asegura que ϕ tiene un máximo global en la frontera de la región factible.

3.3. Normalidad asintótica.

Uno de los objetivos fundamentales del presente estudio es establecer el comportamiento asintótico de la solución al problema replanteado de superficies de respuesta. En la siguiente sección se presentan los resultados que fundamentan el hecho de que si X_0 es el óptimo a tal problema, éste se distribuye asintóticamente normal. Con tal propósito considere los siguientes resultados preliminares.

Lema 3.1 Sean

$$\hat{\theta} = \begin{bmatrix} \hat{\beta} \\ \hat{\sigma}^2 \end{bmatrix}$$

estimadores insesgados (Apéndice B) de

$$\theta = \begin{bmatrix} \beta \\ \sigma^2 \end{bmatrix}$$

entonces $\hat{\theta} \xrightarrow{d} N(\theta, \Sigma)$ donde

$$\Sigma = \begin{bmatrix} \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} & 0 \\ 0 & 2(\sigma^2)^2 / (n-p) \end{bmatrix}$$

y \mathbf{X} es la matriz de diseño.

Prueba. Se sabe que $\hat{\beta}$ se distribuye normal sólo resta determinar la distribución de $\hat{\sigma}^2$, como

$$\frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-p}^2 \quad \text{y} \quad \chi_{n-p}^2 \xrightarrow{d} N[(n-p), 2(n-p)]$$

se tiene que la varianza asintótica de $\hat{\sigma}^2$ está dada por

$$\text{Var}(\hat{\sigma}^2) = \frac{2(\sigma^2)^2}{n-p}$$

además se tiene que $\hat{\sigma}^2$ es independiente (estocásticamente) de $\hat{\beta}$ luego $\text{Cov}(\hat{\sigma}^2, \hat{\beta}_i) = 0 \quad \forall i = 1, 2, \dots, p$, donde $p = 1 + 2k + C_2^k = \text{rango de la matriz de diseño}$, así

$$\hat{\theta} \xrightarrow{-d} N(\theta, \Sigma)$$

Teorema 3.3. Considere el siguiente problema de optimización

$$\text{Min}_{\theta} \phi(\hat{\theta}, r; \theta) = Z^T(x) \hat{\beta} - r \left[\frac{1}{g_1} + \frac{1}{g_2} + \frac{1}{g_3} \right]$$

con

$$g_1 = (\hat{\beta} - \beta)^T X^T X (\hat{\beta} - \beta) - p \hat{\sigma}^2 F \leq 0$$

$$g_2 = \frac{(n-p) \hat{\sigma}^2}{\chi_1} - \sigma^2 \leq 0$$

$$g_3 = \sigma^2 - \frac{(n-p) \hat{\sigma}^2}{\chi_2} \leq 0$$

Asuma que:

i) Las siguientes derivadas

$$\frac{\partial^2 \phi(\circ)}{\partial \theta \partial \theta^T} \quad \text{y} \quad \frac{\partial^2 \phi(\circ)}{\partial \theta \partial \hat{\theta}^T}$$

existen y son continuas en la vecindad del punto $[\hat{\theta}(\theta), \theta]$.

ii) $\hat{\theta} \xrightarrow{-d} N(\theta, \Sigma)$ (lema 3.1).

Entonces el mínimo $\hat{\theta}_0(\hat{\theta})$ es tal que

$$\sqrt{n} [\hat{\theta}_0(\hat{\theta}) - \theta_0(\theta)] \xrightarrow{-d} N(0, V1)$$

o alternativamente

$$\hat{\theta}_0(\hat{\theta}) \xrightarrow{d} N\left[\theta_0(\theta), \frac{1}{n}V_1\right]$$

donde

$$V_1 = \left[\frac{\partial \hat{\theta}_0(\theta)}{\partial \hat{\theta}} \right] \sum \left[\frac{\partial \hat{\theta}_0(\theta)}{\partial \hat{\theta}} \right]^T$$

Prueba. El punto óptimo $\hat{\theta}_0(\hat{\theta})$ se obtiene al solucionar el siguiente sistema de ecuaciones

$$\frac{\partial \phi(\hat{\theta}, r; \theta)}{\partial \theta} = 0$$

equivalentemente

$$\begin{bmatrix} Z^T(x) - r[2X^T X (\hat{\beta} - \beta) g_1^{-2}] \\ -r[g_2^{-2} - g_3^{-2}] \end{bmatrix} = 0$$

sabemos por ii) que

$$\hat{\theta} \xrightarrow{d} N(\theta, \Sigma)$$

por el teorema 2.13

$$\sqrt{n}[\hat{\theta}_0(\hat{\theta}) - \theta_0(\theta)] \xrightarrow{d} N(0, V_1)$$

donde

$$v_1 = \left[\frac{\partial \hat{\theta}_0(\theta)}{\partial \hat{\theta}} \right] \sum \left[\frac{\partial \hat{\theta}_0(\theta)}{\partial \hat{\theta}} \right]^T$$

la forma explícita de

$$\left[\frac{\partial \hat{\theta}_0(\theta)}{\partial \hat{\theta}} \right]$$

se encuentra apoyándose en el teorema de la función implícita (Apéndice A), así

$$\left[\frac{\partial \hat{\theta}_0(\theta)}{\partial \hat{\theta}} \right] = - \left[\frac{\partial^2 \phi(\hat{\theta}_0, r | \theta)}{\partial \theta \partial \theta^T} \right]^{-1} \left[\frac{\partial^2 \phi(\hat{\theta}_0, r | \theta)}{\partial \theta \partial \hat{\theta}^T} \right]$$

donde

$$\frac{\partial^2 \phi(\hat{\theta}_0, r | \theta)}{\partial \theta \partial \theta^T} = \begin{bmatrix} 2r[X^T X][g_1 I_{s-4}(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T X^T X]g_1^{-3} & 0 \\ 0 & -2r[g_2^{-3} + g_3^{-3}] \end{bmatrix}$$

y

$$\frac{\partial^2 \phi(\hat{\theta}_0, r | \theta)}{\partial \theta \partial \hat{\theta}^T}$$

es igual a

$$\begin{bmatrix} -2r[X^T X][g_1 I_{s-4}(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T X^T X]g_1^{-3} & -4rpFX^T X(\hat{\beta} - \beta)g_1^{-3} \\ 0 & -2r(n-p)[g_2^{-3}\chi_1^{-1} + g_3^{-3}\chi_2^{-1}] \end{bmatrix}$$

para invertir la matriz

$$\frac{\partial^2 \phi(\hat{\theta}_0, r|\theta)}{\partial \theta \partial \theta^T}$$

utilizemos el teorema A.9. del apéndice. Así

$$\left[\frac{\partial^2 \phi(\hat{\theta}_0, r|\theta)}{\partial \theta \partial \theta^T} \right]^{-1} = \begin{bmatrix} [2r[X^T X][g_1 I_s - 4(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T X^T X]g_1^{-3}]^{-1} & 0 \\ 0 & -[2r]^{-1} [(g_2 g_3)^3 / g_2^3 + g_3^3] \end{bmatrix}$$

por último

$$\frac{\partial \hat{\theta}_0(\theta)}{\partial \hat{\theta}} = \begin{bmatrix} -I_s - 2pF\{g_1 I_s - 4(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T X^T X\}^{-1}(\hat{\beta} - \beta) \\ 0_{-(n-p)} [(g_2 g_3)^3 / g_2^3 + g_3^3][g_2^{-3} \chi_1^{-1} + g_3^{-3} \chi_2^{-1}] \end{bmatrix}$$

utilizando el teorema A.7 del apéndice.

$$\{g I_s - 4(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T X^T X\}^{-1}$$

tome

$$A = g_1 I_s$$

$$U = -4(\hat{\beta} - \beta)$$

$$V^T = (\hat{\beta} - \beta)^T X^T X$$

así

$$\{ \circ \}^{-1} = g_1^{-1} \left[I_s + \left\{ 4(\hat{\beta} - \beta) (\hat{\beta} - \beta)^T X^T X \right\} / \left\{ -5g_1 + 4p\sigma^2 F \right\} \right]$$

y por lo tanto

$$\frac{\partial \hat{\theta}_0(\theta)}{\partial \hat{\theta}} = \begin{bmatrix} I_s & 2pF \left[I_s + \left\{ 4(\hat{\beta} - \beta) (\beta - \hat{\beta}) X^T X^T \right\} / \left\{ -5g_1 + 4p\sigma^2 F \right\} \right] (\hat{\beta} - \beta) \\ 0 & (n-p) \left[(g_2 g_3)^3 / g_2^3 + g_3^3 \right] \left[g_2^{-3} \chi_1^{-1} + g_3^{-3} \chi_2^{-1} \right] \end{bmatrix} \circ$$

Lema 3.2. Sea $Z^T(x)$ como en (2.3), entonces

$$\frac{\partial Z^T(x)}{\partial x} = M$$

donde

$$M = [0_k \mid I_k \mid 2\text{Diag}(X) \mid C]$$

con

$$C = [C_1 \mid C_2 \mid \dots \mid C_{k-1}]$$

$$C_i \in \mathbb{R}^{k \times k-1}$$

$$C_i = \begin{bmatrix} 0_1 \\ 0_2 \\ \vdots \\ 0_{i-1} \\ X^T A_i^T \\ X^T I_{i-1} \end{bmatrix}$$

$0_i \in \mathbb{R}^{1 \times k-i}$ si $i = 0$ no existe en la columna

$$A_i \in \mathbb{R}^{k-i \times k}$$

$$A_i = \begin{bmatrix} 0_1 & \cdots & 0 & e_i^T \\ & & & e_{k-i}^T \end{bmatrix}$$

$$0_i \in \mathbb{R}^{k-i}$$

$$e_i \in \mathbb{R}^{k-i}$$

$$x^T = [x_1, x_2, \dots, x_k]$$

Prueba. Veamos la forma de la derivada

$$\frac{\partial Z^T(x)}{\partial x} = \begin{bmatrix} \frac{\partial Z^T(x)}{\partial x_1} \\ \frac{\partial Z^T(x)}{\partial x_2} \\ \vdots \\ \frac{\partial Z^T(x)}{\partial x_k} \end{bmatrix}$$

$$\text{donde } Z^T(x) = \left[1, x_1, \dots, x_k, x_1^2, \dots, x_k^2, x_1 x_2, \dots, x_1 x_k, \dots, x_{k-1} x_k \right]$$

derivando término a término se puede ver que

$$\frac{\partial Z^T(x)}{\partial x} = \begin{bmatrix} 0 & | & 1 & 0 \dots 0 & | & 2x_1 & 0 \dots 0 & | & x_2 & x_3 \dots x_k & | & 0 & 0 \dots 0 \dots 0 \\ 0 & | & 0 & 1 \dots 0 & | & 0 & 2x_2 \dots 0 & | & x_1 & 0 \dots 0 & | & x_3 & x_4 \dots x_k \dots 0 \\ \vdots & & & \vdots & \\ \vdots & & & \vdots & \\ 0 & | & 0 & 0 \dots 1 & | & 0 & 0 \dots 2x_k & | & 0 & 0 \dots x_1 & | & 0 & 0 \dots x_2 \dots x_{k-1} \\ \hline 0_k & & I_k & & & & 2\text{diag}(X) & & & & & & C \end{bmatrix}$$

donde C es como antes.

Teorema 3.4. Sea el problema

$$\text{Max}_x \phi(x, r | \theta_0) = Z^T(x) \beta_0 + r \left[\frac{1}{g} \right]$$

donde

$$g = x^T x - C^2 \leq 0$$

Suponga además:

i) Las siguientes derivadas

$$\frac{\partial^2 \phi(\circ)}{\partial \theta_0 \partial \theta_0^T} \quad \text{y} \quad \frac{\partial^2 \phi(\circ)}{\partial \theta_0 \partial \hat{\theta}_0^T}$$

existen y son continuas en la vecindad del punto $[\hat{x}(\theta_0), \theta_0]$.

$$\text{ii) } \hat{\theta}_0 \xrightarrow{d} N \left[\theta_0(\theta), \frac{1}{n} V_1 \right] \text{ (Teorema 3.3).}$$

Entonces el máximo es tal que

$$\hat{x}(\hat{\theta}_0) \xrightarrow{d} N \left[x(\theta_0), \frac{1}{n} V_2 \right]$$

donde

$$V_2 = \left[\frac{\partial \hat{x}(\theta_0)}{\partial \hat{\theta}_0} \right] V_1 \left[\frac{\partial \hat{x}(\theta_0)}{\partial \hat{\theta}_0} \right]^T$$

y

$$\frac{\partial \hat{x}(\theta_0)}{\partial \hat{\theta}_0} = \left[2B - 2r \left\{ gI - 4X^T X / g^3 \right\} \right]^{-1} \left[M | O_k \right]$$

Siendo M como en el lema anterior y B como en 2.4.

La prueba a este teorema es totalmente análoga a la del teorema 3.3.

Para concluir este capítulo cabe hacer algunas observaciones a los resultados anteriores.

1) Con el propósito de que estos resultados sean aplicables, estimadores consistentes de tales funciones paramétricas pueden ser propuestos (teorema 2.10) y por el teorema 2.14 las densidades encontradas en los teoremas 3.3 y 3.4 pueden ser sustituidas por las correspondientes estimaciones.

2) Observe además que los estimadores insesgados de varianza uniformemente mínima (EIVUM) empleados en los teoremas 3.3 y 3.4 ($\hat{\theta}$) son consistentes ya que son una combinación lineal de estimadores de verosimilitud máxima, lo anterior se desprende del teorema 2.10. Finalmente como los óptimos $\hat{\theta}_0(\theta)$ y $x(\theta_0)$ son funciones de los EIVUM se concluye que estos son estimadores consistentes, esto es

$$\hat{\theta}_0(\theta) \xrightarrow{P} \theta_0(\theta) \text{ y}$$

$$\hat{x}(\hat{\theta}_0) \xrightarrow{P} x(\theta_0)$$

4. LITERATURA CITADA

- Armitano, O., E. Edelman y U. García P. 1985. Programación no lineal. Ed. Limusa. México. 216 p.
- Cooper, L., and D. Steinberg. 1970. Introduction to methods of optimization. W.B. Saunders Company. United States of America. 381 p.
- Cramer, H. 1970. Métodos matemáticos de estadística. Cuarta Edición. Ed. Aguilar. España. 660 p.
- Charnes A. and W.W. Cooper. 1959. Chance constrained programing. Management Science. Vol. 6: 73-79.
- Dupačová, J. 1984. Stability in stochastic programing with recourse-estimated parameters. Mathematical programing. 28 (1):72-83. North Holland. United States of America
- Fiacco, A.V., and G. P. McCormick. 1966. Extensions for SUMT for nonlinear programing: equality constraints and extrapolation. Management Science. 12 (11): 816-828.
- Graybill, F.A. 1969. Introduction to matrices with applications in Statistics. Wadsworth Publishing Company. United States of America. 382 p.
- Graybill, F.A. 1976. Theory and application of the Linear Model. Duxbury Press. United States of America. 704 p.
- Khuri, A.I. and J.A. Cornell. 1987. Response Surfaces. Designs and Analyses. Marcel Dekker, Inc. United States of America. 405 p.

- Melaku, A. 1986. Asymptotic normality of the optimal allocation in multivariate stratified random sampling. *Sankhyá: The Indian Journal of Statistics*. 48, Series B, Pt. 2, 224-232 p.
- Mood, A.M., F.A., Graybill and D.C., Boes. 1983. *Introduction to the theory of Statistics*. Third Edition. McGraw-Hill International Book Company. United States of America. 564 p.
- Myers, R.H. 1971. *Response surface methodology*. Allyn and Bacon, Inc. United States of America. 246 p.
- Prawda W., J. 1981. *Métodos y modelos de Investigación de operaciones*. Vol. 2 Modelos estocásticos. Ed. Limusa. México. 1026 p
- Prawda W., J. 1982. *Métodos y Modelos de Investigación de operaciones*. Vol. 1 Modelos determinísticos. Ed. Limusa. México. 935 p.
- Prékopa, A. 1978. *The use of stochastic programming for the solution of some problems in statistics and probability*. Technical Summary report. University of Mathematics Research Center. 19p.
- Rao, R. 1973. *Linear Statistical inference and its applications*. Second edition. John Wiley and sons. United States of America. 625 p.
- Rao, S.S. 1979. *Optimization theory and applications*. Wiley Eastern. India. 711 p.
- Rudin, W. 1980. *Principios de Análisis matemático*. Tercer Edición. Ed. McGraw-Hill. México. 370p.
- Srivastara. M.S. and C.G. Khatri. 1979. *An Introduction to multivariate Statistics*. Elsevier North Holland, Inc. United States of America. 350 p.
- Vivanco T., J.C. 1988. *Fundamentos del Análisis de las superficies de respuesta*. Tesis. Maestría en Ciencias. Universidad Autónoma Agraria "Antonio Narro". Buenavista, Saltillo, Coah., México. 113 p.

APENDICES

Apéndice A. ALGEBRA DE MATRICES

Definición A.1. (Inverso de una matriz). Sea A una matriz cuadrada. Si existe una matriz B tal que $AB = BA = I$, entonces a B se le llama inverso de A , denotado por A^{-1} y a la matriz A se le conoce como no singular. En caso contrario se dice que es singular.

Teorema A.1. Si una matriz tiene inverso, este es único.

Teorema A.2. Si A tiene inverso, entonces A^{-1} tiene inverso y $(A^{-1})^{-1} = A$.

Teorema A.3. Si A y B son matrices no singulares, entonces AB tiene un inverso y $(AB)^{-1} = B^{-1}A^{-1}$. Este resultado puede extenderse a cualquier número finito de matrices.

Definición A.2. (Transpuesta de una matriz). Si los renglones y columnas de una matriz A son intercambiados, la matriz resultante es conocida como transpuesta y se denota como A^T . Si A es de tamaño $m \times n$, entonces A^T es de dimensión $n \times m$.

Teorema A.4. Si A es una matriz, entonces $(A^T)^T = A$.

Teorema A.5. Sean dos matrices cualesquiera A y B tal que el producto AB se encuentra definido; entonces

$$(AB)^T = B^T A^T$$

Definición A.3. (Matriz Simétrica). Si $A = A^T$, entonces a A se le conoce como matriz simétrica.

Teorema A.6. Sea A una matriz cualquiera, entonces $A^T A$ y AA^T son simétricas.

Teorema A.7. Sea A una matriz no singular, y U y V dos vectores columna, entonces

$$(A + UV^T)^{-1} = A^{-1} - \frac{(A^{-1}U)(V^T A^{-1})}{1 + V^T A^{-1}U}$$

Definición A.4. (Inverso Generalizado). Sea A una matriz $m \times n$. Si existe una matriz A^- tal que satisface las siguientes cuatro condiciones, entonces A^- es uninverso generalizado de A.

1. AA^- es simétrica.
 2. A^-A es simétrica.
 3. $AA^-A = A$.
 4. $A^-AA^- = A^-$.
- (A.1)

Teorema A.8. Para cada matriz A existe una y solo una matriz A^- que satisface las condiciones de la ecuación A.1.

Definición A.5. (Matrices particionadas). En ocasiones es conveniente el representar una matriz por medio de la yuxtaposición de dos o más matrices en forma particionada. Así una matriz particionada se representa como

$$A = \begin{pmatrix} P & Q \\ R & S \end{pmatrix}$$

donde no hay posibilidad de confusión debido a que el número de renglones en P es igual al que existe en Q, y las

columnas de P iguales a las de R.

Teorema A.9. Si A y D son matrices simétricas tales que los inversos que surran en la siguiente expresión existan, entonces:

$$\begin{pmatrix} A & B \\ B^T & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + FE^{-1}F^T & -FE^{-1} \\ -E^{-1}F^T & E^{-1} \end{pmatrix}$$

donde $E = D - B^T A^{-1} B$ y $F = A^{-1} B$.

Definición A.6. (Matriz semidefinida positiva) Una matriz A $n \times n$ se define como semidefinida positiva si y solo si

1. $A = A^T$
2. $y^T A y \geq 0$ para todo vector $y \in \mathbb{R}^n$ tal que $y \neq 0$ y la igualdad se cumple al menos se cumple para un vector y

Definición A.7. (Matriz definida positiva) Una matriz A $n \times n$ se define como semidefinida positiva si y solo si

1. $A = A^T$
2. $y^T A y > 0$ para todo vector $y \in \mathbb{R}^n$ tal que $y \neq 0$.

Definición A.8. (Matriz no negativa). Una matriz se define como no negativa si y solo si esta es positiva definida o semidefinida positiva.

Teorema A.10. Sea $X \in \mathbb{R}^n$ y $f: \mathbb{R}^n \rightarrow \mathbb{R}$ entonces si

i) $f(X) = a^T X$ con $a \in \mathbb{R}^n$ constante

$$\frac{\partial f(X)}{\partial X} = a$$

ii) $f(X) = X^T A X$ con A simétrica, constante de orden n
entonces

$$\frac{\partial f(X)}{\partial X} = 2AX$$

(Graybill, 1969 y Rao, 1973)

Para finalizar este apéndice se enuncia el siguiente resultado llamado Teorema de la Función Implícita.

Teorema A.11. Sea $f: \mathbb{R}^{m+n} \rightarrow \mathbb{R}^m$, tal que $f(x_0, y_0) = 0$ para algún punto $(x_0, y_0) \in \mathbb{R}^{m+n}$. Suponga además que

$$\left. \frac{\partial f(x, y)}{\partial x} \right|_{x_0, y_0}$$

es no singular. Entonces existe $A \in \mathbb{R}^{m+n}$ y $B \in \mathbb{R}^m$, con $(x_0, y_0) \in A$ y $y_0 \in B$, tales que :

i) para todo $y \in B$ corresponde un único x tal que

$$(x, y) \in A \quad \text{y} \quad f(x, y) = 0$$

ii) Si x es tal que $x = g(y)$, entonces $g: B \in \mathbb{R}^m \rightarrow \mathbb{R}^n$, con $g(y_0) = x_0$,

$$f(g(y), y) = 0, \quad y \in B$$

$$\frac{\partial g(y)}{\partial y} = - \left[\frac{\partial f(x_0, y_0)}{\partial x} \right]^{-1} \left[\frac{\partial f(x_0, y_0)}{\partial y} \right]$$

(Rudin, 1980)

columnas de P igual en número que las de R.

Teorema A.9. Si A y D son matrices simétricas tales que los inversos que ocurran en la siguiente expresión existan, entonces:

$$\begin{pmatrix} A & B \\ B^T & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + FE^{-1}F^T & -FE^{-1} \\ -E^{-1}F^T & E^{-1} \end{pmatrix}$$

donde $E = D - B^T A^{-1} B$ y $F = A^{-1} B$.

Definición A.6. (Matriz semidefinida positiva) Una matriz A n x n se define como semidefinida positiva si y solo si

1. $A = A^T$
2. $y^T A y \geq 0$ para todo vector $y \in \mathbb{R}^n$ tal que $y \neq 0$ y la igualdad se cumple al menos se cumple para un vector y

Definición A.7. (Matriz definida positiva) Una matriz A n x n se define como semidefinida positiva si y solo si

1. $A = A^T$
2. $y^T A y > 0$ para todo vector $y \in \mathbb{R}^n$ tal que $y \neq 0$.

Definición A.8. (Matriz no negativa). Una matriz se define como no negativa si y solo si esta es positiva definida o semidefinida positiva.

Teorema A.10. Sea $X \in \mathbb{R}^n$ y $f: \mathbb{R}^n \rightarrow \mathbb{R}$ entonces si

i) $f(X) = a^T X$ con $a \in \mathbb{R}^n$ constante

$$\frac{\partial f(X)}{\partial X} = a$$

ii) $f(X) = X^TAX$ con A simétrica, constante de orden n entonces

$$\frac{\partial f(X)}{\partial X} = 2AX$$

(Graybill, 1969 y Rao, 1973)

Para finalizar este apéndice se enuncia el siguiente resultado llamado Teorema de la Función Implícita.

Teorema A.11. sea $f: \mathbb{R}^{m+n} \rightarrow \mathbb{R}^m$, tal que $f(x_0, y_0) = 0$ para algún punto $(x_0, y_0) \in \mathbb{R}^{m+n}$. Suponga además que

$$\left. \frac{\partial f(x, y)}{\partial X} \right|_{x_0, y_0}$$

es no singular. Entonces existe $A \in \mathbb{R}^{m+n}$ y $B \in \mathbb{R}^m$, con $(x_0, y_0) \in A$ y $y_0 \in B$, tales que :

i) para todo $y \in B$ corresponde un único x tal que

$$(x, y) \in A \quad \text{y} \quad f(x, y) = 0$$

ii) Si x es tal que $x = g(y)$, entonces $g: B \in \mathbb{R}^m \rightarrow \mathbb{R}^n$, con $g(y_0) = x_0$,

$$f(g(y), y) = 0, \quad y \in B \quad \text{y}$$

$$\frac{\partial g(y)}{\partial y} = - \left[\frac{\partial f(x_0, y_0)}{\partial x} \right]^{-1} \left[\frac{\partial f(x_0, y_0)}{\partial y} \right]$$

(Rudin, 1980)

Apéndice B. REGRESION

B.1. Modelo Lineal General de Regresión.

Sea Y un vector de variables aleatorias observables $n \times 1$; Sea X una matriz $n \times p$ de constantes; Sea β un vector $p \times 1$ de parámetros desconocidos; Además sea ε un vector aleatorio $n \times 1$, más aún ε se distribuye $N(0, \sigma^2 I)$. Si estas cantidades se relacionan por medio de la expresión:

$$Y = X\beta + \varepsilon \quad (B.1)$$

definen al modelo lineal general. (Graybill, 1976)

En forma analoga el modelo se puede expresar como:

$$Y \sim N(X\beta, \sigma^2 I) \quad (B.2)$$

B.1.1. Estimación de β y σ^2 .

Los estimadores insesgados de varianza uniformemente mínima para β y σ^2 están dados por las expresiones:

$$\hat{\beta} = (X^T X)^{-1} X^T Y = X^{-} Y$$

$$\tilde{\sigma}^2 = \frac{1}{n} Y^T (I - XX^T) Y$$

Más aún $\hat{\beta}$ se distribuye normal debido a que es una combinación lineal de un vector normal, además:

$$E(\hat{\beta}) = E(X^T Y) = X^T E(Y) = X^T X \beta = (X^T X)^{-1} X^T X \beta = \beta$$

$$\begin{aligned} \text{Cov}(\hat{\beta}) &= \text{Cov}(X^T Y) = X^T \text{Cov}(Y) X^{-T} \\ &= X^T \sigma^2 I X^{-T} \\ &= \sigma^2 X^T X^{-T} \\ &= \sigma^2 (X^T X)^{-1} \end{aligned}$$

concluyendo

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1}) \quad (\text{B.3})$$

Por otra parte se sabe que:

$$\frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-p}^2$$

y que $\hat{\sigma}^2$ es un estimador insesgado de σ^2 . Además

$$\hat{\sigma}^2 = \frac{n}{n-p} \tilde{\sigma}^2$$

donde $\tilde{\sigma}^2$ es el estimador de máxima verosimilitud de σ^2 .

Para demostrar que $\hat{\beta}$ y $\hat{\sigma}^2$ son independientes recuerde el siguiente teorema:

Teorema B.1. Sea Y un vector aleatorio $n \times 1$ distribuido $N(\mu, \Sigma)$, donde Σ tiene rango n . Si $B\Sigma A = 0$, la forma cuadrática $Y^T A Y$ es independiente de las formas

lineales BY donde B es una matriz $q \times n$.

Se sabe que:

$$\hat{\beta} = X^{-1}Y$$

$$n\sigma^2 = Y^T(I - XX^{-1})Y$$

$$Y \sim N(X\beta, \sigma^2 I)$$

$$\text{tomemos } B = X^{-1} \quad \Sigma = \sigma^2 I \quad A = (I - XX^{-1})$$

así:

$$\begin{aligned} B\Sigma A &= X^{-1}\sigma^2 I (I - XX^{-1}) \\ &= \sigma^2 I (X^{-1} - X^{-1}XX^{-1}) \\ &= 0 \end{aligned}$$

se concluye, por lo tanto, independencia de $\hat{\beta}$ y $\hat{\sigma}^2$.

B.1.2. Regiones de confianza para $\hat{\beta}$ y $\hat{\sigma}^2$.

De las ecuaciones B.2 y B.3 se tiene que :

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1}) \quad \text{y} \quad \frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-p}^2$$

Sea A tal que $X^T X = AA^T$ pues $X^T X > 0$, luego

$$\frac{1}{\sigma} (\hat{\beta} - \beta)^T A \sim N(0, I)$$

por lo tanto, como $AA^T = X^T X$

$$\frac{(\hat{\beta} - \beta)^T X^T X (\hat{\beta} - \beta)}{\sigma^2} \sim \chi_p^2$$

luego

$$\frac{\frac{(\hat{\beta} - \beta)^T X^T X (\hat{\beta} - \beta)}{p \hat{\sigma}^2}}{\frac{(n-p) \hat{\sigma}^2}{(n-p) \sigma^2}} = \frac{(\hat{\beta} - \beta)^T X^T X (\hat{\beta} - \beta)}{p \hat{\sigma}^2} \sim F_{p, n-p}$$

entonces el elipsoide

$$(\hat{\beta} - \beta)^T X^T X (\hat{\beta} - \beta) \leq p \hat{\sigma}^2 F_{p, n-p} \quad (\text{B.4})$$

define una región de confianza para β .

Por otro lado dado que:

$$\frac{(n-p) \hat{\sigma}^2}{\sigma^2} \sim \chi_{n-p}^2$$

y usando esta como cantidad pivote, se puede construir $I\sigma$, la cual define un intervalo de confianza para σ

$$I\sigma = \left\{ \frac{(n-p) \hat{\sigma}^2}{\chi_{\alpha/2, n-p}^2} \leq \sigma^2 \leq \frac{(n-p) \hat{\sigma}^2}{\chi_{1-\alpha/2, n-p}^2} \right\} \quad (\text{B.5})$$

con coeficiente de confianza $1 - \alpha$. (Graybill, 1976)

Dado que $\hat{\sigma}^2$ y $\hat{\beta}$ son independientes una región de confianza simultanea para ambos tendría la estructura de las ecuaciones (B.5) y (B.4) con una probabilidad conjunta igual al producto de las individuales. (Mood, Graybill y Boes, 1983).

Finalmente suponga que se desea estimar un modelo de la forma

$$Y(x) = Z^T(x)\beta + \varepsilon$$

donde

$$Z^T(x) = \left[1, x_1, \dots, x_k, x_1^2, \dots, x_k^2, x_1 x_2, \dots, x_1 x_k, \dots, x_{k-1} x_k \right]$$

con tal efecto solo defina la matriz X en la ecuación (B.1) tal que las columnas correspondan a las etiquetas definidas por el vector $Z^T(x)$. Así todos los resultados anteriores son aplicables a este modelo, solo note que en este caso $p = 1 + 2k + C_2^k$.