



# **Agraria**

**Nueva Epoca**

Año I · Volumen 1 · Número 1 · Enero - Abril de 2004 · Buenavista, Saltillo, Coah., México



***Número especial de Estadística***

**Teoría General Sobre Listado de un Conjunto**

***Pag. 8***

**Caracterización y Aplicación de la Distribución  
Birnbbaum-Sanders**

***Pag. 16***



# 4<sup>o</sup> UNIVERSIDAD AUTÓNOMA AGRARIA ANTONIO NARRO SIMPOSIO NACIONAL DE HORTICULTURA



## DISEÑO



## MANEJO



## PRODUCCIÓN



CONFERENCIAS SOBRE:  
Diseño de Invernaderos  
Construcción  
Manejo de Sustratos  
Nutrición  
Clima Controlado  
Manejo de Problemas Fitosanitarios  
Cultivos Susceptibles de Explotación  
Manejo de Hortalzas  
Manejo de Ornamentales  
Otros Cultivos  
Importancia Nacional de los Invernaderos  
Competitividad  
Concurso de Carteles

## INSTITUCIONES PARTICIPANTES

Departamento de Horticultura  
Universidad Autónoma Agraria Antonio Narro  
Buenavista, Saltillo, Coahuila.

Departamento de Horticultura  
Universidad Autónoma Agraria Antonio Narro  
Unidad Laguna  
Torreón, Coahuila.

Instituto Tecnológico  
Agropecuario N° 10  
Ejido Aha, Torreón, Coahuila.

Facultad de Agricultura y Zootecnia de la  
Universidad Juárez del Estado de Durango  
Ejido Venecia, Gómez Palacio, Durango.

13, 14 y 15 DE OCTUBRE DE 2004

SEDE: CENTRO DE CONVENCIONES DE TORREÓN, COAHUILA, MÉXICO.

Paseo de la Rosita No. 908, Col. Campestre La Rosita, CP 27250

INVERNADEROS



## Directorio

**Dr. Adalberto Benavides Mendoza**

Director de Investigación

**Dr. Andres Martínez Cano**

Subdirector de Programación y Evaluación

**Ing. Pedro Recio del Bosque**

Subdirector de Operación de Proyectos

## Unidad Laguna

**Dr. Esteban Favela Chávez**

Subdirector de Investigación

## Comité Editorial

**Dr. Miguel A. Capó Arteaga**

Editor en Jefe

**Dr. Jesús Valdés Reyna**

Editor Ejecutivo

## Diseño y Formación

Miguel A. Estrada Villarreal

## Secretario de Producción

M. Ed. Víctor M. López González

## Colaboradores

M. C. Ricardo Cuéllar Flores

M. C. Cecilia Burciaga

## Editores Técnicos

**Dr. José L. Puente Manriquez**

Fitomejoramiento, UL

**Dr. Raúl Rodríguez García**

Riego y Drenaje

**Dr. Jesús M. Fuentes Rodríguez**

Produccion Animal



La **Revista Agraria -Nueva Epoca-** es una publicación científica, cuatrimestral, de la UAAAN.

Domicilio Conocido, Buenavista, Saltillo, Coah., México

<http://www.uaaan.mx/DirInv/portalagraria/portal.htm> · [agraria\\_ne@uaaan.mx](mailto:agraria_ne@uaaan.mx)

Tel. (844) 411-02-00, Ext. 2404 . Fax 411-02-11



**Centeotl**, deidad azteca de la agricultura, es una advocación de Chicomecóatl, diosa del maíz. La UAAAN, en su afán de rescatar los valores del pasado histórico de México la ha adoptado como logotipo de su revista científica, como símbolo que evoca y reafirma nuestras raíces culturales

**NUESTRA PORTADA.** Composición de fachadas -izquierda a derecha- de la Dirección de Investigación y de la Rectoría de la UAAAN, abajo, un paisaje desértico..

## Contenido

Mensaje del Rector / <i>Message of the Rector</i>	1
Normas Editoriales / <i>Instructions for authors</i>	4
Presentación / <i>Presentation</i>	5
Convocatoria/ <i>Paper call</i>	6
Artículos / <i>Articles</i>	
Propiedades Estadísticas del Muestreo por Línea Intercepto y Cuadros Cargados en la Estimación de la Cobertura y Densidad Vegetales . <i>Dino Ulises González Uribe, Félix de Jesús Sánchez Pérez</i>	7
<i>Statistics Properties of the 'Line-Intercept' and 'Stocked-Quadrat' Sampling in the Estimation of Plant Cover and Density</i>	
Teoría General sobre el Listado de un Conjunto. <i>Gerardo Sánchez Martínez</i>	11
<i>General Theory on the Members of a Set''''</i>	
Caracterización y Aplicación de la Distribución Birnbaum-Saunders como Modelo de Tiempos de Vida . <i>Edgar Espinosa Cuello, Mario Cantú Sifuentes, Víctor Leiva Sánchez</i>	19
<i>Characterization and Application of the Birnbaum-Saunders Distribution as a Life Time Model</i>	
Índice de Funcionamiento de Normas de Paro en Control Estadístico de Calidad . <i>Juan Manuel Saucedo Esquivel</i>	29
<i>Performance Index of Stopping Rules in Statistical Control of Quality</i>	
Una Aplicación de los Algoritmos Genéticos en la Discriminación <i>Aurora Montano Rivas, Mario Cantú Sifuentes, Nidia Hernández Pérez</i>	42
<i>An Application of Genetic Algorithms in Discrimination</i>	

## Mensaje del Rector

---

La actividad científica sólo se considera completa cuando los resultados obtenidos se publican en un medio que cuenta con arbitraje por pares académicos, cuando los resultados se transfieren al sector productivo en forma de una variedad registrada, una patente, un desarrollo tecnológico, o cuando se vende, o un reporte técnico es utilizado por la industria.

Como parte de nuestras acciones para lograr una presencia cada vez mayor de la Universidad ante la sociedad, el sector agropecuario y forestal y la comunidad científica nacional e internacional, surge la revista Agraria -Nueva Epoca- para dar continuidad a los esfuerzos editoriales iniciados con Agraria en 1985. La Universidad se justifica en la medida en que es útil, y una de las formas de aumentar su presencia, importancia y rentabilidad es la rápida divulgación y transferencia de los resultados de la investigación básica y aplicada que se produce a través de los proyectos de investigación.

Es compromiso institucional apoyar a Agraria -Nueva Epoca- de tal manera que se asegure su calidad y continuidad, así como su adecuada distribución. A largo plazo esperamos que Agraria -Nueva Epoca- se constituya en un proyecto financieramente sostenible que se ubique en los índices nacionales e internacionales. Para lograrlo se requiere no sólo el compromiso institucional; también es necesaria la participación decidida de los profesores investigadores y de los estudiantes mediante el envío de sus colaboraciones a nuestra revista científica. Recordemos que este proyecto, al igual que muchos otros. es un esfuerzo de toda la Comunidad Universitaria.



ALMA TERRA MATER  
Dr. Luis Alberto Aguirre Uribe  
Rector

## Normas Editoriales

### Tipo de materiales para publicación

La revista **Agraria**-Nueva Epoca- acepta, para su publicación, materiales en español e inglés, sobre temas relacionados con las ciencias agrícolas, pecuarias y forestales, incluyendo las áreas de ingeniería, agroindustria y socioeconómicas. Todo material deberá venir acompañado de la solicitud correspondiente.

Estos materiales pueden ser artículos científicos, notas de investigación o ensayos.

Los materiales que se envíen para su publicación deberán ceñirse a las normas que, para tal efecto establece **Agraria**-Nueva Epoca- y estarán sujetos a revisión y arbitraje por el Comité Editorial de la revista -o por quienes éste designe-, como requisito previo a su publicación.

No se aceptan trabajos ya publicados, o que estén sometidos a consideración en otros medios científicos de difusión.

Es de desear que la realización de la investigación, cuyos materiales sean enviados para su publicación, no exceda de 4 años anteriores a la fecha de su remisión.

### Formato

El respeto a las siguientes indicaciones respecto al formato solicitado facilitará grandemente nuestro trabajo de edición.

### Textos

Los textos, con todos sus anexos, deberán enviarse empaquetados (nosotros preferimos WinZip), sin contraseñas de seguridad, por correo electrónico, escritos en un procesador de textos de uso común (preferimos Word), en formato tamaño carta (21.57 x 27.94 cm), sin sangría, y a renglón seguido, con márgenes de 2.5 cm por lado. Agradeceremos evitar nombres de archivo excesivamente largos o con espacios en blanco. Los textos se redactarán en un tipo formal conocido ttf (True Type Font) tales como Arial, Times New Roman o similares, de 12 puntos. Las notas se escribirán en 9 puntos.

Todos los renglones, incluidos los encabezados, se iniciarán, invariablemente, a partir del margen izquierdo, sin sangría.

Todos los encabezados, independientemente de su orden, se escribirán en altas y bajas, y negrillas.

Los párrafos se escribirán sin pasar renglón entre ellos; para separarlos, a fin de hacer el texto fácil de leer y corregir, se utilizará el formato automático de párrafo del procesador, para darles un espaciado posterior de 6 puntos.

Las palabras no se separarán, en ningún caso, por sílabas. Es conveniente desactivar el comando automático de inserción de guiones (hyphenation) de su procesador.

El material no deberá exceder de 520 líneas para artículos científicos y ensayos, y de 200 líneas para una nota científica, incluidos cuadros y figuras.

Las unidades que se empleen serán las del Sistema Internacional de Unidades (<http://www.sc.ehu.es/sbweb/fisica/unidades/unidades/unidades.htm>)

Las páginas, al igual que los cuadros y las figuras, se numerarán progresivamente con números arábigos.

**Cuadros y figuras.** Los cuadros y las figuras contendrán sólo la información esencial y en ningún caso repetirán los datos que se presenten en el texto, o en otra forma. Cuadros y figuras deben ser claros, simples, concisos e ilustrativos.

Los cuadros no excederán, en ningún caso, los márgenes de impresión arriba mencionados y deberán presentarse en el cuerpo del texto, con el formato correspondiente, con las columnas separadas por tabulaciones, sin espacios a mano, y en la posición en que se espera que aparezcan, con el número de orden correspondiente.

En los cuadros se empleará sólo el número de cifras significativas necesarias para destacar el punto que se desee.

Los cuadros se realizarán en formato básico con tres líneas horizontales continuas: al inicio del cuadro, al inicio del cuerpo del cuadro (no en el encabezamiento) y al final. El campo y el encabezamiento de las columnas se pueden dividir a conveniencia del autor. No se deben añadir líneas verticales. Los encabezamientos, de columnas y líneas, se escribirán con minúsculas, excepto la primera letra de la oración. Las unidades se colocan debajo de la segunda línea horizontal, como en el ejemplo que se proporciona.

Las figuras tampoco excederán, en ningún caso, los márgenes de impresión establecidos. La posición que vaya a ocupar cada figura, deberá estar indicada en el texto con negrillas, en renglón aparte, con el número correspondiente.

Cada figura se enviará en archivo por separado, en formato tif (compresión LZW), o jpg, con el tamaño exacto en que se pretende que aparezca en la publicación, en una resolución no inferior a 150 pixeles por pulgada, con el número que le corresponda (p. ej: fig 01.jpg).

Los puntos experimentales deberán marcarse visiblemente. Para dividir los ejes, se escogerán intervalos constantes para cada uno. Los mosaicos fotográficos deberán entregarse montados en un solo archivo gráfico (tif, o jpg), totalmente terminados. El aumento de las microfotografías debe indicarse en la leyenda.

En archivo por separado se enviará un listado de las figuras incluidas en el material enviado, con el número de orden y el pie de grabado correspondientes (p. ej.: listafigs.doc)

Las figuras pueden ser fotos a color o en tonos de gris -según sea su original-, gráficas (de preferencia a color), ilustraciones, dibujos, o grabados (de preferencia a color).

Los cuadros deberán redactarse en el mismo procesador de textos y formato señalado arriba.

Las ecuaciones, si las hubiera, se insertarán en el texto con un editor de ecuaciones compatible con su procesador.

### Notas de pie de página

Sólo se podrán utilizar, cuando sean absolutamente indispensables, para identificar información adicional y se numerarán progresivamente en el texto. Los asteriscos se reservarán para indicar significancia a 5% (\*) y 1% (\*\*), respectivamente. En el pie de grabado -o de cuadro- se incluirán las notas o llamadas que sean pertinentes, y serán señaladas con números arábigos.

### Citas bibliográficas

Las citas bibliográficas deberán ser de literatura reciente, relevante y sólo las exclusivamente necesarias para sustentar los planteamientos hechos.

Más detalles en <http://www.uaaan.mx/DirInv/Convoc/conv-web/normas.htm>, o <http://www.uaaan.mx/DirInv/portal-agrarialportal.htm>

## Presentación

---

Agraria, Revista Científica de la Universidad Autónoma Agraria Antonio Narro, vio su primer número en marzo de 1985. Desde entonces, funciona como uno de los mecanismos institucionales para la promoción y difusión del quehacer científico y tecnológico de los profesores y estudiantes involucrados en actividades de investigación.

Los cambios en la política pública de apoyo a la ciencia y la tecnología, obligan a nuestra Institución a modificar sus enfoques y proyectos de manera proactiva. Considerando que la actividad editorial y de difusión de la ciencia y la tecnología no quedan exentas de esta evolución, es que surge Agraria -Nueva Epoca- como una publicación especializada en la difusión de temas relacionados con las ciencias agrícolas, pecuarias y forestales, incluyendo lo relativo a los aspectos económicos, sociales, de ingeniería y agroindustrias.

De una revista científica se exige que incluya información relevante, actualizada y útil para la comunidad científica y el sector productivo, se pide, además, que cumpla con una serie de criterios que le permitan acceder a los índices internacionales como el ISI, CabAbstracts, Biological, Abstracts y, a nivel nacional, en el índice de Revistas Mexicanas de Investigación Científica y Tecnológica del CONACYT.

Partiendo de lo anterior y como resultado de una serie de reuniones en donde participaron los representantes de Investigación de la UAAAN en Saltillo y Torreón, el Subdirector de Postgrado y un conjunto de profesores investigadores, se conformó una estructura editorial, con un Comité Editorial y un Consejo Editorial que incluye un Editor, Editores técnicos y Arbitros de la UAAAN y de otras Instituciones; dicha estructura promueve el esfuerzo coordinado y eficaz de las autoridades universitarias y los profesores investigadores junto con sus colaboradores de ésta y otras instituciones. El objetivo final de esta acción es lograr que Agraria -Nueva Epoca- se posicione, a través de un esfuerzo sostenido, en los ya mencionados índices nacionales e internacionales, para darle mayor presencia y actualidad a las disciplinas, proyectos y resultados logrados por nuestra Universidad.

Este primer número de la Revista Agraria Nueva Epoca, es posible gracias a la colaboración del personal que labora en la Dirección de Investigación y en la Dirección de Comunicación, al trabajo del Comité Editorial y del Consejo Editorial, así como a los aportes científicos de los autores y coautores cuyos trabajos aparecen en este número. Por adelantado se agradecen las futuras colaboraciones de investigadores no sólo de nuestra Universidad sino de otras instituciones hermanas.

Para concluir, es necesario mencionar que el Comité Editorial decidió que el primer número de Agraria -Nueva Epoca- se dedique a los artículos enviados por profesores investigadores dedicados principalmente a la investigación y aplicaciones de la Estadística. Esperamos que en el conjunto de números de Agraria -Nueva Epoca- que se publiquen a partir de este momento, algunas de ellas constituyan número de carácter monográfico dedicados a alguna temática relevante y específica que surja del interés de la comunidad científica y tecnológica.

  
Dr. **Adalberto Benavides Mendoza**  
Director de Investigación

## CONVOCATORIA

La Dirección de Investigación de la Universidad Autónoma Agraria Antonio Narro

### CONVOCA

a los investigadores, nacionales y extranjeros, interesados en publicar artículos científicos, a enviar sus contribuciones a la revista **Agraria** -Nueva Epoca-, bajo las siguientes bases:

Los trabajos recibidos se someterán al proceso de revisión descrito en las Normas Editoriales de la Revista.

Se recibirán contribuciones inéditas de todos los interesados, nacionales o extranjeros, en español o inglés.

Los artículos deberán versar sobre temas de contenido agrícola, pecuario, forestal, y socioeconómico del entorno regional, nacional e internacional.

Las modalidades de publicación son las siguientes:

#### Artículo científico

Es el resultado de un trabajo de investigación en el cual se aplicó, de forma rigurosa, el método científico, estudiando el efecto que tienen diferentes tratamientos sobre la respuesta medible de un sistema, como metodología para comprobar o rechazar una hipótesis claramente establecida en el trabajo.

Los artículos científicos que se envíen deberán constar de las siguientes partes: Título, Título en inglés, Autor(es), Institución(es) de adscripción y datos de localización del autor responsable (domicilio, teléfono, fax, e-mail), Abstract, que es la traducción al inglés del Resumen, incluidas las palabras clave; Resumen, que incluirá al pie las palabras clave hasta un máximo de seis, Introducción, Materiales y métodos, Resultados y discusión, Conclusiones, Literatura citada, Agradecimientos.

#### Ensayo científico

Consiste en el análisis crítico de una recopilación actualizada de artículos científicos, informes de investigación, o materiales similares, en los que el autor o autores aportan su opinión personal sobre un tema, estableciendo conclusiones respecto al estado actual del conocimiento sobre el mismo.

Partes de que consta el Ensayo: Título, Título en inglés, Autor(es), Institución(es) de adscripción y datos de localización del autor responsable (domicilio, teléfono, fax, e-mail), Abstract, que es la versión al inglés del Resumen, incluye las palabras clave, Resumen, incluidas las palabras clave hasta un máximo de siete, Introducción, Desarrollo del tema, con los subtítulos que se estimen convenientes, Discusión, cuando proceda, Conclusiones, Literatura citada.

#### Nota de investigación

Son materiales basados en trabajos experimentales que, sin perjuicio del método y rigor científicos, presentan aspectos metodológicos innovadores o resultados que, por su carácter novedoso, el autor considera de interés publicar antes de finalizar su investigación.

La nota, aunque de menor extensión, cubre todos los aspectos relevantes del proceso de investigación. Su estructura es similar a la del artículo científico, y trata cada uno de sus apartados, con menor profundidad y detalle, aunque no tiene que incluir los encabezados.

La excepción a lo anterior son el Abstract, que se omite, y la Literatura citada, apartado que deberá incluirse expresamente.

De ser necesario, podrán incluirse -también- algún cuadro o ilustración, cuando resulten relevantes para la mejor comprensión de la nota.

Los trabajos a publicar deberán hacerse llegar en versión electrónica, acompañados de una solicitud, de conformidad con las especificaciones marcadas en las Normas Editoriales arriba mencionadas, a la siguiente dirección electrónica: *agraria\_ne@uaaan.mx*, con atención a:

Editor en Jefe de la Revista Agraria -Nueva Epoca-

Dirección de Investigación, UAAAN, Domicilio conocido, Buenavista,

Saltillo, Coahuila, México. CP. 25315

Para mayor información respecto a esta Convocatoria visite: <http://www.uaaan.mx/DirInv/portal~agraria/portal.htm> para consultas diríjase al Editor en Jefe: *agraria\_ne@uaaan.mx*.

# Propiedades Estadísticas del Muestreo por Línea Intercepto y Cuadros Cargados en la Estimación de la Cobertura y Densidad Vegetales<sup>1</sup>

Dino Ulises González Uribe<sup>2</sup>

Félix de Jesús Sánchez Pérez<sup>3</sup>

Universidad Autónoma Agraria Antonio Narro, Depto. de Estadística y Cálculo, [www.uaaan.mx](http://www.uaaan.mx)

---

**Abstract.** *Statistics Properties of the 'Line-Intercept' and 'Stocked-Quadrat' Sampling in the Estimation of Plant Cover and Density.* The statistical properties of 'unbiased estimator', sufficiency and consistency for the line intercept, and 'stocked-quadrat' sampling were theoretically studied; in both cases the properties in the variables; canopy cover, and density in units of interest were demonstrated. Relative efficiency was proven taking solely the vegetal density for unbiased and consistent estimators, being the highest the estimation for line intercept.

**Key words:** Canopy cover, density, 'unbiased estimator', 'sufficient e.', 'consistent e.', 'relative efficiency', 'line-intercept' sampling, 'stocked-quadrat' sampling.

**Resumen.** Se estudiaron teóricamente las propiedades estadísticas de insesgamiento, suficiencia y consistencia para el muestreo por línea intercepto y cuadros cargados; en ambos casos se demostraron esas propiedades para las variables cobertura y densidad en unidades de interés. Tomando únicamente a la densidad vegetal para estimadores insesgados y consistentes se probó la eficiencia relativa, resultando superior la estimación por línea intercepto.

**Palabras clave:** Cobertura vegetal, Densidad vegetal, Estimador insesgado, Estimador suficiente, Estimador cconsistente, Eficiencia relativa, Muestreo por línea intercepto, Muestreo por cuadros cargados.

---

## Introducción

En los inventarios de vegetación, frecuentemente se utiliza el muestreo estadístico para obtener información rápida, veraz y económica para la toma de decisiones. El objetivo del muestreo, en este caso, es la obtención de una estimación descriptiva de algunas características de la población vegetal en estudio, como lo son la cobertura y la densidad (Burguete y Carrillo, 1972; Lyon, 1968). Esta estimación debe representar suficientemente el parámetro en estudio y permitir detectar con precisión las diferencias entre poblaciones vegetales (Lyon, 1968).

Al muestreo que utiliza líneas rectas para conocer cobertura y densidad se le llama muestreo por **línea intercepto**; se le denomina así por considerar en la

evaluación a aquellos individuos que se cortan por la línea en su parte aérea. Se utiliza porque es de fácil aplicación. Si son cuadros, se cuentan sólo aquellas unidades de muestreo con los individuos de interés para obtener así la densidad vegetal; a este procedimiento se le denomina muestreo por **cuadros cargados**. El rango de aplicación de ambos procedimientos de muestreo es muy amplio (Cochran, 1950; Kaiser, 1983; Swindel, 1983).

Dado que el interés es la estimación de un parámetro de una característica, como la media poblacional de la cobertura y/o la media poblacional de la densidad a partir de una muestra, la estimación está sujeta a riesgo, entre otras razones, debido a la estructura del estimador con la que se infiere el valor del parámetro. Por tal motivo, en un diseño de muestreo se propone un estimador y se analizan sus propiedades como el sesgo, eficiencia, consistencia y otras de relevancia que son señaladas en teoría estadística, como la suficiencia (Burguete y Carrillo, 1972).

---

<sup>1</sup> Título de la Tesis presentada por el autor para obtener el grado de Maestría en Ciencias en Estadística Experimental

<sup>2</sup> Autor

<sup>3</sup> Coautor y Asesor Principal de la Tesis

Si el estimador utilizado para calcular la media poblacional de la cobertura y la media poblacional de la densidad vegetal posee el mayor número de estas propiedades deseables, entonces se considera de buena **calidad**, por lo tanto la estimación de ambas variables es satisfactoria y, sin duda alguna, se puede utilizar en la estimación de ambos parámetros (Burguete y Carrillo, 1972; Kisinger et al., 1960). Dada la utilización de los procedimientos de muestreo mencionados, en este estudio se propone como objetivo demostrar las propiedades estadísticas básicas de los estimadores de la línea intercepto y muestreo por cuadros cargados, los cuales son: insesgamiento, suficiencia, consistencia y eficiencia.

### Materiales y Métodos

#### Descripción del Muestreo por Línea Intercepto

El uso de la línea intercepto puede definirse como un procedimiento de muestreo de vegetación basado en la medición de todas las plantas interceptadas por un plano vertical de líneas, localizadas aleatoriamente y de igual longitud (Canfield, 1941). Aunque también puede hacerse la estimación con líneas de diferente longitud (McDonald, 1980; Butler y McDonald, 1983). Con el muestreo por línea intercepto puede determinarse la cobertura de corona y la densidad vegetal (Figura 2.1).

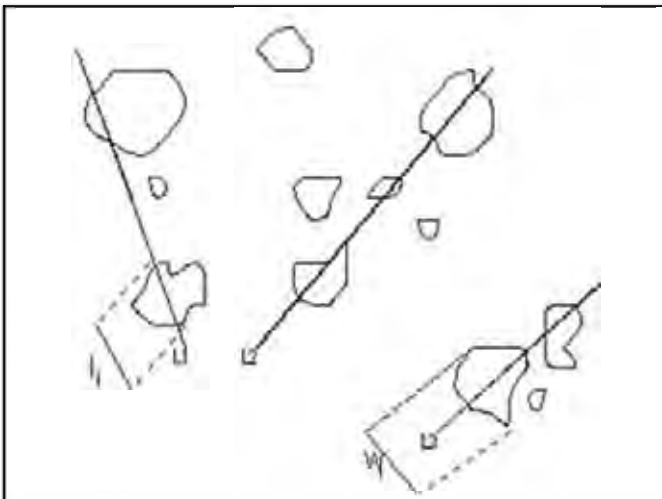


Figura 2.1 Área de estudio con  $M= 12$  individuos de interés y  $n= 3$  unidades de muestreo

#### Descripción del Muestreo por Cuadros Cargados

Si tenemos un área **A** que se subdivide en  $N$  unidades de muestreo en forma de cuadro, cada uno de ellos de área **a**, a los cuadros con la presencia de individuos de interés serán los cuadros cargados (Figura 2.2). Si se denota por  $\theta$  el número de cuadros no cargados en una

muestra de unidades de muestreo de tamaño **n**, se puede obtener el número de individuos en el área **A**.

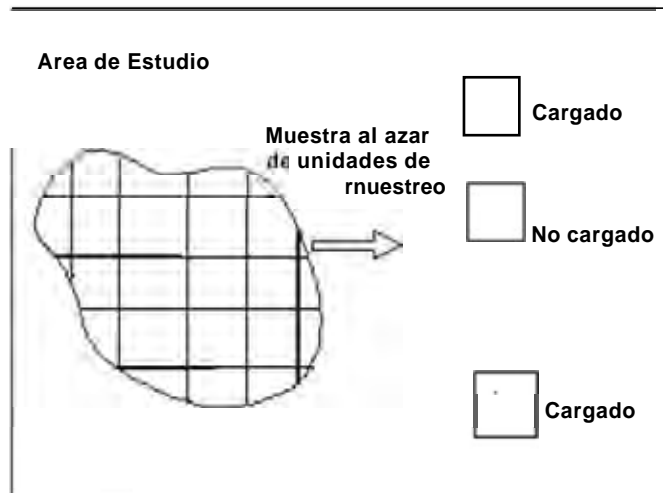


Figura 2.2 Cuadros cargado y no cargado tomados de un área **A**

### Estimación y Estimador

Un estimador es una fórmula, la cual establece cómo calcular un valor dado contenido en una muestra aleatoria que se obtiene en campo; un estimador se designa como  $\hat{\theta}$  y se toma como si fuera el valor verdadero de una población al cual se le llama parámetro; el parámetro  $\theta$  sólo se conocerá si se realiza un censo de la población; por esta razón, el estimador es de gran importancia en el muestreo. La acción de utilizar al estimador y conocer las consecuencias de utilizarlo como una función de decisión al tomar el valor del estimador como si fuera el parámetro, es la estimación.

Los estimadores de la cobertura y densidad vegetal en el muestreo por línea intercepto y muestreo por cuadros cargados, estiman a la media poblacional del parámetro, y si el estimador usado posee la propiedad de que su valor esperado, o esperanza matemática sea igual al parámetro se dice que el estimador es insesgado, teóricamente, si

$$E(\hat{\theta}) = \theta.$$

Como la media poblacional se estima, este valor varía de acuerdo con la muestra aleatoria que se tome en campo; si se conoce su varianza se puede estimar su variación con respecto a su media. La consistencia de un estimador se prueba en la varianza, de la media poblacional cuando el tamaño de muestra crece; si al crecer la varianza se hace cero, se dice que la varianza es consistente y ha alcanzado su máxima eficiencia.

También es de interés saber si la muestra aleatoria contiene la información necesaria para estimar el parámetro; cuando esto sucede, se dice que el estimador

posee la propiedad de suficiencia, para lo cual se necesita saber la función de distribución de la variable en un estudio de población, aunque se puede suponer, no obstante que la distribución normal es la de uso más frecuente (Burguete y Carrillo, 1972)

Cuando dos o más estimadores insesgados estiman a la misma media de la población, puede escogerse para su uso aquél estimador que tenga la menor varianza, a lo cual se le llama eficiencia relativa.

### Resultados y Discusión

El estimador de la media poblacional de la cobertura vegetal que se usa en el muestreo por línea intercepto es insesgado. Si se quiere estimar la cobertura de corona de alguna especie vegetal, a partir de una muestra aleatoria de  $n$  unidades de muestreo en una población, el estimador a usar junto con su varianza es:

$$\hat{c} = 25\pi \left( \frac{\sum_{i=1}^m l_i}{\sum_{j=1}^n L_j} \right)$$

$$\hat{V}(\hat{c}) = \frac{s_l^2}{nL^2}$$

$$s_l^2 = \frac{1}{n-1} \left( \sum_{i=1}^n l_i - \hat{c}L \right)^2$$

Si es una cobertura cuadrada o rectangular, se utiliza la expresión dentro del paréntesis (para el caso de la media poblacional); si la varianza estimada de la media de la cobertura es consistente, el estimador de la media también es eficiente.

Si el estimador de la densidad vegetal en el muestreo por línea intercepto estima insesgadamente a la media poblacional, su varianza es consistente y, además, el estimador posee la propiedad de suficiencia. Los estimadores encontrados son los siguientes:

$$U = \frac{\sum_{i=1}^m w_i}{\sum_{j=1}^n L_j} * area$$

$$\hat{V}(U) = \frac{s_w^2}{nL^2}$$

$$s_w^2 = \frac{1}{n-1} \left( \sum_{i=1}^n w_i^2 - \frac{n\hat{D}}{A} \right)$$

En el muestreo por línea intercepto  $L$  la suma total de las unidades de muestreo  $l_i$  y  $w_i$  son el intercepto y ancho máximo, respectivamente, de las coberturas interceptadas (ver Figura 2.1).

El estimador de la media poblacional de la densidad vegetal y varianza del muestreo por cuadros cargados, posee las propiedades deseables de estimación; en este caso, la función de distribución Poisson es la adecuada para el muestreo. Así entonces, se contarán en aquella muestra de  $n$  unidades como 0, a los que no tienen presencia de individuos de interés o no cargados, y como 1, a los que sí tienen individuos de interés o cargados (ver Figura 2.2):

$$\hat{D} = -\left(\frac{1}{a}\right) \ln\left(\frac{y}{n}\right)$$

$$\hat{\tau} = N\hat{D}$$

$$\hat{V}(\hat{D}) = \frac{e^{\hat{D}a} - 1}{na^2}$$

Para obtener estimaciones de la media de la población de la densidad vegetal con dos estimadores insesgados, es mejor utilizar el estimador del muestreo por línea intercepto, ya que su varianza es muy pequeña comparada con la del muestreo por cuadros cargados.

### Conclusiones

Conocer las propiedades estadísticas de los estimadores que se usan en un trabajo de muestreo, hace que estemos más confiados en las estimaciones que obtendremos con ellos. No basta saber que el estimador es insesgado. Su varianza y consistencia son importantes, puesto que no sólo se conocerá una estimación puntual, sino que también se podrá construir un intervalo de confianza para la media poblacional y se podrá inferir, con menor error, sobre los parámetros de la población.

En el muestreo por línea intercepto, es necesario conocer la forma promedio de la corona del individuo de interés, para saber qué estimador utilizar; el estimador de la media poblacional de la cobertura junto con el de densidad, hacen que este procedimiento de muestreo proporcione más información sobre una población, por lo que su uso es recomendable.

El muestreo por cuadros cargados proporciona, con rapidez, la estimación de la media poblacional de la densidad vegetal; sin embargo, es muy alta con respecto a la de la línea intercepto, por lo que es recomendable un estudio más profundo sobre este procedimiento de muestreo para conocer con amplitud sus aplicaciones.

**Literatura Citada**

- 1 Burguete, H. J. F. y A. C. Liz. 1972. Algunas propiedades de los estimadores en muestreo por áreas. *Agrociencia*. No. 10. 91-104.
- 2 Butler, A. S. and L. L. McDonald. 1983. Unbiased systematic sampling plan for the line intercept method. *J. Range Management*. 36(4): 463-468.
- 3 Canfield, H. R. 1941. Application of the line interception method in sampling range vegetation. *J. of Forestry*. 388-394.
- 4 Cochran, G. W. 1950. Estimation of bacterial densities by means of the "most probable number". *Biometrics*. 105-115.
- 5 Kaiser, L. 1983. Unbiased estimation in line-intercept sampling. *Biometrics*. 39: 965-976.
- 6 Kisinger, E. F.; R. E. Eckert and P. O. Currie. 1960. A comparison of the line-interception, variable plot and loop methods as used to measure shrub-crown cover. *J. Range Management*. 13: 17-21.
- 7 Lyon, J. 1968. An evaluation of density sampling methods in a shrub community. *J. Range Management*. 21: 16-20.
- 8 McDonald, L. L. 1980. Line-intercept sampling for attributes other than coverage and density. *J. Wildlife Management*. 44(2): 530-533.
- 9 Swindel, F. B. 1983. Choice of size and number of quadrats to estimate density from frequency in poisson and binomially dispersed populations. *Biometrics*. 39: 455-464.

# Teoría General sobre el Listado de un Conjunto

Gerardo Sánchez Martínez

Depto. de Estadística y Cálculo, Universidad Autónoma Agraria Antonio Narro, Buenavista, Saltillo, Coah., México

---

**Abstract.** *General Theory on the listing of a set. The aim of this note is to focus on the problem of producing a list of all members of a set. The main purpose is the construction of an order relationship in the studied set. A general algorithm for the listing must be developed, which has a great importance for the calculation of estimators using nonparametric methods.*

**Key Words:** Permutation, Order Relation, Successor of an Element, Minimum, Maximum, Lexicographical Order, Algorithm.

**Resumen.** El propósito de este trabajo es dar un enfoque al problema de predecir una lista de todos los miembros de un conjunto. La intención principal es la construcción de una relación de orden en el conjunto de interés. Se elaborará un algoritmo general para el listado, el cual es de gran importancia para el cálculo de estimadores usando métodos no paramétricos.

**Palabras clave:** Permutación, Relación de orden, Sucesor de un elemento, Mínimo, Máximo, Orden lexicográfico, Algoritmo.

---

## Introducción

En este trabajo se analiza el siguiente problema general de enumeración o listado: Dado un conjunto finito  $A$ , producir una lista en la que cada elemento de  $A$  aparezca solo una vez. A primera vista, éste parece ser un problema simple; después de todo, una manera de especificar un conjunto es precisamente proporcionando una lista de todos los elementos. Por ejemplo, al escribir  $A = \{3, 6, 9, 12, 15\}$ , se están enumerando los miembros de  $A$ . Sin embargo, existe otra forma de determinar un conjunto, la cual se denomina método de comprensión o regla, de acuerdo a la cual los elementos de conjunto se distinguen mediante una propiedad que los caracteriza; vea por ejemplo, Suples (1971), o Dolciani et al. (1986). Por ejemplo, si  $B$  es el conjunto de todos los enteros positivos menores de 17 que son múltiplos de tres, se puede establecer con toda seguridad que (a)  $2 \notin B$ , pues a pesar de que dos es un entero positivo y menor a 17, dos no es múltiplo de tres, (b)  $21 \notin B$ , pues aunque 21 si es un entero positivo múltiplo de tres, 21 no es menor a 17, y (c)  $9 \in B$ , pues nueve es un entero positivo que es tanto múltiplo de tres como menor a 17. El aspecto fundamental en esta discusión es que la propiedad que se estipula para los miembros de  $B$  permite decir, sin ambigüedad alguna, si un objeto determinado pertenece o no a dicho conjunto. Aun cuando originalmente se especifique un conjunto mediante una caracterización de sus miembros, frecuentemente es necesario disponer de una lista de

elementos, esto es, debe utilizarse la propiedad que distingue a los objetos que constituyen  $B$  para producir una lista de todos sus miembros— equivalentemente, para enumerar todos los elementos de  $B$ . En el caso especificado en el párrafo precedente, no es difícil mostrar que  $B = \{3, 6, 9, 12, 15\}$ . Es decir  $B$  coincide con  $A$ , de manera que en estas circunstancias es sencillo pasar de la caracterización del conjunto  $B$  por el método de comprensión, a la descripción de  $B$  mediante el listado o enumeración de sus miembros. Sin embargo, en general la situación es más compleja. Considere, por ejemplo, el conjunto  $C$  que consiste de todos los dígitos (enteros entre cero y nueve, inclusive), que aparecen por lo menos cuatro ocasiones dentro de las primeras veinte cifras de la parte fraccionaria en la expansión decimal del número  $\pi$ . En este caso,  $C$  consta de a lo más cinco miembros; producir una lista de  $C$  sin embargo, requiere de que se tenga la expansión de  $\pi$  hasta 20 cifras significativas, y es necesario de disponer de algoritmos especiales para este fin. Este ejemplo muestra que, a pesar de que un conjunto tiene pocos miembros— a lo más cinco en caso de  $C$ — generar una lista de sus elementos no es, en todos los casos, un problema simple; más bien, con frecuencia reclama un análisis esmerado y, tal vez, complejo. En el caso del conjunto  $C$  puede mostrarse que consiste de un único elemento, a saber, cinco (Papadopoulos (1997)).

El propósito general de este trabajo es diseñar un procedimiento para enumerar los miembros de un conjunto

finito arbitrario, y para alcanzar este objetivo, la expansión ha sido organizada de la manera siguiente: En la Sección 2 se discute la relevancia del problema de listado en el campo de la estadística no paramétrica, tratando de ubicar el problema general abordado desde una óptica que muestre su relevancia. Posteriormente, en la Sección 3 se discute la relación entre el problema general de listado, y la construcción de un orden total en el conjunto de interés, y se introducen las nociones de elemento máximo y mínimo en un conjunto ordenado. La Sección 4 trata sobre la idea de sucesor de un elemento, la cual es una noción central para formular, en la Sección 5, el resultado principal de este trabajo, a saber, el algoritmo general de listado. Después de establecer este procedimiento, en la Sección 6 se introduce una relación de orden total entre vectores de  $R^n$ , a saber, el denominado orden lexicográfico (Munkres (1975)). y la exposición concluye con algunos comentarios breves en la Sección 7.

### El Angulo Estadístico

Antes de emprender la tarea de diseñar un método para enumerar los miembros de un conjunto finito arbitrario, es conveniente enfatizar, una vez más, la razón por la cual este problema es de gran importancia en el análisis estadístico. El punto de partida es la observación de que enumerar (listar) los elementos de un conjunto finito puede ser un problema complejo por dos razones:

- (a) Es difícil determinar explícitamente los elementos del conjunto, por ejemplo, este en el caso para el conjunto  $C$  considerado en la Sección 1.
- (b) El conjunto de interés tiene un número grande de elementos.

En el análisis estadístico, particularmente al emplear métodos de estimación no paramétrica, la dificultad para enumerar los miembros de un conjunto de interés se origina, frecuentemente, por la segunda de las causas mencionadas. Por ejemplo, considere las siguientes tres clases de objetos construidos a partir de una población  $P$  que consta de  $N$  objetos, los cuales, sin pérdida de generalidad alguna, se supone que son los primeros  $N$  enteros positivos, i.e.,

$$P = \{1, 2, \dots, N\} \tag{2.1}$$

- La clase  $P(N, N)$ , la cual consiste de todas las permutaciones de tamaño  $N$  de la población  $P$ .

$$P(N, N) = \left\{ (a_1, a_2, \dots, a_n) \mid a_i \in P, i=1, 2, \dots, n, a_i \neq a_j, 1 \leq i \neq j \leq N \right\} \tag{2.2}$$

### Relaciones de Orden

La observación clave para abordar el problema de listar un conjunto finito  $A$ , es que cuando se produce una lista de sus miembros, de inmediato se induce un 'orden' en el conjunto, donde los elementos que se escriben primero son 'menores' que aquéllos que aparecen después. Para aclarar esta idea, considere el conjunto

$$A = \{A, *, 0, !\} \tag{3.1}$$

Suponga ahora que se pide producir una lista de los miembros de  $A$ . Ante este requerimiento, se podría responder con

$$!, A, *, 0 \tag{3.2}$$

En este caso, si se convienen en que un elemento es menor que cualquiera de los que están a su derecha en la lista, se tendría que

$$\begin{aligned} '!' \text{ es menor que } 'A' \quad '!' \text{ es menor que } '*' \text{ y } '!' \text{ es menor que } '0' \\ 'A' \text{ es menor que } '*', \quad 'A' \text{ es menor que } '0' \quad \text{y} \quad '*' \text{ es menor que } '0' \end{aligned} \tag{3.3}$$

Esta 'relación de orden' fue inducida por el listado (3.2), y es claro que al formar otra lista de los miembros de  $A$ , la relación de orden se alterará. Por ejemplo, si en vez de (3.2) se enumeran los miembros de  $A$  como

$$A, !, *, 0$$

Entonces la primera afirmación en (3.3) debe de cambiarse a 'A' es menor que '!'. La discusión puede extenderse a cualquier conjunto, en el sentido de que al disponer de una lista de miembros, puede definirse una relación de orden. El propósito de esta sección es establecer la proposición inversa: Si se dispone de una relación de orden adecuada en un conjunto finito A, entonces, dicha relación puede utilizarse para generar un listado de A, el resultado se demostrará posteriormente en el Teorema 3.1. Por el momento, es oportuno definir, de manera precisa, cual es la idea de relación de orden que se utilizará.

**Definición 3.1.** [Munkres (1975), Suples (1971)]. Sea A un conjunto finito, y considere una relación, denotada mediante  $\prec$  definida en el conjunto. Esta relación es un orden completo en A si satisface los siguientes requerimientos:

(a) [Transitividad] Para  $a, b, c \in A$ ,

$$a \prec b \text{ y } b \prec c \Rightarrow a \prec c.$$

(b) [Completez] Para todos los miembros  $a, b \in A$ , se tiene que exactamente una de las siguientes alternativas es valida.

$$a \prec b, a = b, b \prec a.$$

En esta definición, la condición (a) es bastante natural desde un punto de vista intuitivo, mientras que el requerimiento (b) impone la condición de que dos elementos distintos de A sean comparables, i.e., si  $a \neq b$ , debe tenerse que  $a \prec b$ , o bien  $b \prec a$ . Es costumbre leer la expresión  $a \prec b$  como 'a precede a b', en lugar de 'a es menor a b'; el propósito es no confundir la relación de orden recientemente introducida, con el orden usual en los números reales cuando  $A = \mathbb{R}$ . Por otro lado, el hecho de que a preceda a b, esto es,  $a \prec b$ , también se escribe como  $b \succ a$ , esto es,

$$a \prec b \Leftrightarrow b \succ a. \quad (3.4)$$

La expresión  $b \succ a$  se lee como 'b sucede a a', o 'b es posterior a a'; similarmente  $a \prec b$  con frecuencia se expresa verbalmente como 'a es anterior a b'.

**Ejemplo 3.1.** Sea  $A = \{A, *, 0, !\}$  el conjunto considerado al inicio de esta sección. La relación (3.3) puede expresarse, con la notación recientemente introducida, como

$$! \prec A, ! \prec *, ! \prec 0, A \prec *, A \prec 0 \text{ y } * \prec 0 \quad (3.5)$$

Con esta definición de la relación de precedencia, no es difícil verificar que las propiedades de transitividad y

completez se satisfacen. Por lo tanto (3.5) determina un orden completo en A.

Dado un conjunto finito A dotado con un orden completo  $\prec$ , existen dos elementos especiales, a saber, el elemento mínimo y el elemento máximo.

**Definición 3.2.** Sea A un conjunto finito dotado con un orden completo  $\prec$ . En este caso, (a)  $m \in A$  es un elemento mínimo si para todo  $a \in A$  con  $a \neq m$ , se tiene que  $m \prec a$ . Similarmente, (b)  $M \in A$  es un elemento máximo si siempre que  $a \in A$  sea tal que  $a \neq M$ , entonces  $a \prec M$ .

**Teorema 3.1.** Sea A un conjunto finito y no vacío dotado de un orden completo. En este caso, (a) A posee un elemento mínimo, Más aún, (b) A tiene exactamente un elemento mínimo. Similarmente, (c) A posee un elemento máximo, y (d) El elemento máximo de A es único.

**Demostración.** (a) El argumento es por inducción en el número de elementos de A, el cual se denota por n. Primero observe que si A tiene  $n = 1$  elementos, entonces la afirmación es clara. Suponga que si un conjunto dotado con un orden completo tiene k miembros, entonces tiene un elemento mínimo, y sea A un conjunto de  $n = k + 1$  elementos. En estas circunstancias, se verificará que A posee un elemento mínimo. Con esta finalidad, seleccione un elemento cualquiera,  $a^* \in A$ , y defina conjunto  $\bar{A}$  mediante

$$\bar{A} = A \setminus \{a^*\} \quad (3.6)$$

En este caso,  $\bar{A}$  es un conjunto de k elementos, y por lo tanto tiene un miembro mínimo, por la hipótesis de inducción. Sea  $\bar{m} \in \bar{A}$  un miembro mínimo de  $\bar{A}$ , i.e.,

$$a \in \bar{A} \text{ y } a \neq \bar{m} \Rightarrow \bar{m} \prec a. \quad (3.7)$$

Para demostrar que el conjunto original tiene un mínimo, primero observe que  $\bar{m} \in \bar{A}$ , y por lo tanto  $\bar{m} \neq a^*$  (vea(3.6)). Considerando que dos elementos distintos de A son comparables, por definición 3.1, esto implica que ocurre alguna de las siguientes dos alternativas:

$$a^* \prec \bar{m} \text{ o } \bar{m} \prec a^*$$

**Caso 1:**  $a^* \prec \bar{m}$ . En estas situación se verificará que  $a^*$  es un elemento mínimo de A, i.e.,

$$a^* \prec a, \text{ si } a \in A \text{ y } a \neq a^*$$

En efecto, si  $a \in A$  es diferente de  $a^*$ , entonces  $a \in \bar{A}$  (vea (3.6)). Por lo tanto,  $a \neq m$ , o bien  $m \prec a$ ; bajo la primera alternativa se tiene que  $a^* \prec a$ , pues se supone que  $a^* \prec m$  y  $m \prec a$ , mientras que si la segunda posibilidad ocurre, se obtiene que  $a^* \prec m$  y  $m \prec a$ , de tal forma que usando la transitividad de la relación de orden, se obtiene  $a^* \prec a$ . Este argumento muestra que la implicación (3.8) es válida, y por lo tanto,  $a^*$  es un elemento mínimo de  $A$ .

**Caso 2:**  $m \prec a^*$  en estas circunstancias,  $m$  es un miembro mínimo de  $A$ . En efecto, seleccione  $a \in A$  tal que  $a \neq m$  y observe que existen dos alternativas posibles: (1)  $a = a^*$ , circunstancia en la que se tiene

$$m \prec a \tag{3.9}$$

pues  $a = a^*$ , y (2) si  $a \neq a^*$ , entonces  $a \in \bar{A}$  (vea(3.6)), y como  $m$  es un mínimo de  $A$ , se tiene que (3.9) también ocurre cuando  $a \neq m$ . En resumen: Suponiendo que un conjunto con k miembros siempre tiene un elemento mínimo, se ha mostrado que cualquier conjunto k+1 elementos también posee un mínimo, completando el argumento que establece la parte (a) por el método de inducción.

(a) Suponiendo que  $m$  y  $m_1$  son dos elementos mínimos de  $A$ , se demostrará que  $m = m_1$ . Primero, observe que

$$m \in A \text{ y } a \neq m \Rightarrow m \prec a \tag{3.10}$$

(Vea la Definición 3.2), y similarmente

$$a \in A \text{ y } a \neq m_1 \Rightarrow m_1 \prec a \tag{3.11}$$

Suponga ahora que  $m \neq m_1$ . En este caso, a partir de(3.10) con  $a = m_1$ , se desprende que

$$m \prec m_1$$

Mientras que utilizando (3.11) con  $a = m$ , se obtiene que

$$m_1 \prec m$$

Sin embargo, las dos últimas desigualdades desplegadas ocurren de forma simultánea, contradiciendo la condición de completez en la Definición 3.1. Esta contradicción muestra que el supuesto del que se origina, i.e.,  $m \neq m_1$ , es falso, de manera que  $m = m_1$ , completando la demostración de la parte (b). El resto del teorema puede demostrarse mediante argumentos análogos a los utilizados para establecer las partes (a) y (b).

### El Sucesor de un Elemento

En esta sección se presenta una importante idea para la formulación de un algoritmo general de listado de los

miembros de un conjunto. En forma simple, esta noción es la de 'elemento siguiente', o 'sucesor inmediato' a un miembro de un conjunto ordenado.

**Definición 4.1.** Sea  $A$  un conjunto finito dotado de un orden completo. Dado  $a \in A$ , sea  $b \in A$  un elemento que satisface las siguientes condiciones:

$$a \prec b \tag{4.1}$$

Y

Si  $c \in A$  satisface  $a \prec c$ , entonces  $b \prec = c$  (4.2)

En este caso,  $b$  se denomina sucesor de  $a$ ; note que en (4.2),  $a \prec b$  es una forma abreviada de la expresión  $a \prec b$  o  $a = b$ :

$$a \prec b \Leftrightarrow a < b \text{ o } a = b$$

Para entender esta idea, suponga, que los elementos de  $A$  se colocan en una línea horizontal, de manera que si  $x, y \in A$  son tales que  $x \prec y$ , entonces  $x$  se encuentra a la izquierda de  $y$ . Suponga ahora que se recorren los miembros de  $A$  de izquierda a derecha; en este caso, un sucesor de  $a$  es un elemento  $b \in A$  que se ubica a la derecha de  $a$  con la propiedad de que al moverse de  $a$  hacia  $b$  no se encontrará elemento alguno de  $A$  'en el camino'. El siguiente teorema muestra que, con la excepción del máximo, todos los miembros de  $A$  tienen un sucesor, y que éste es único.

**Teorema 4.1.** Sea  $A$  un conjunto finito dotado con un orden completo, y suponga que  $a \in A$  es diferente del máximo elemento de  $A$ . En este caso,  $a$  posee un sucesor, el cual es único.

**Demostración.** Considere el conjunto

$$S(a) = \{c \in A \mid a \prec c\} \tag{4.1}$$

esto es,  $S(a)$  contiene a todos los miembros de  $A$  que suceden al elemento  $a$ . Debido a que  $a \neq M$ , se tiene que  $a \prec M$ , de tal forma que  $M \in S(a)$ . Por lo tanto,  $S(a)$  es un conjunto no vacío, el cual posee un único elemento mínimo, por el Teorema 3.1, el cual se denotará mediante  $h$ . Puesto que  $b \in S(a)$ , se tiene que

$$a \prec b$$

vea (4.1), Ahora considere un elemento arbitrario  $c \in A$  tal que  $a \prec c$ . En este caso, se tiene que  $c \in S(a)$ , de manera que  $b \prec = c$ , pues  $h$  es el mínimo de los elementos de  $S(a)$ . Esta discusión puede resumirse como sigue: Se ha demostrado que  $a \prec b$  y que  $b \prec = c$

para todo  $c$  que satisface  $a \prec c$ . Por lo tanto,  $b$  es un sucesor de  $a$ . Para concluir, suponga que  $b$  y  $b$ , son dos sucesores de  $a$ . En este caso,

$$a \prec b \text{ y } a \prec c \Rightarrow b \prec c \quad (4.2)$$

Similarmente

$$a \prec b \text{ y } a \prec c \Rightarrow b \prec c \quad (4.3)$$

Usando la implicación en (4.2) con  $c=b$ , se desprende que

$$b = b,$$

mientras que a partir de la implicación en (4.3) con  $c=b$ , se obtiene

$$b = b,$$

y por lo tanto, si  $b \neq b$ , las dos últimas relaciones desplegadas implican que  $b \prec b$  y  $b \prec b$  ocurren simultáneamente, lo cual se opone a la propiedad de completéz en la Definición de orden total. Por lo tanto,  $b = b$ , estableciendo la unicidad del sucesor de  $A$ .

**Definición 4.1.** Suponga que  $A$  es un conjunto finito que consiste de más de un elemento. Defina la función

$$Suc: A \setminus \{M\} \rightarrow A$$

mediante

$$Suc(a) = \text{sucesor de } a, \quad a \in A \setminus \{M\}.$$

**Ejemplo 4.1.** Considere el conjunto  $A$  introducido en el Ejemplo 3.1, eso es,  $A = \{A, *, 0, !\}$ ; en el cual la relación de orden completo está dada por

$$! \prec A, ! \prec *, ! \prec 0, A \prec 0, \text{ y } * \prec 0 \quad (4.4)$$

Como se mencionó anteriormente,  $M = 0$  es el máximo de  $A$ . El sucesor de cada elemento de  $A$  diferente de 0 está dado, de acuerdo a (4.4), por

$$Suc(!) = A, \quad Suc(A) = *, \quad Suc(*) = 0. \quad (4.5)$$

Como consecuencia del Teorema 4.1 y de la definición de la función **Suc**, se desprende el siguiente corolario.

**Corolario 4.1.** Dado un conjunto finito con más de un elemento, la función **Suc** es uno a uno. Más aún, **Suc** es una biyección entre  $A \setminus \{M\}$  y  $A \setminus \{m\}$ .

**Demostración.** Primero se verificará que **Suc** es una función uno a uno, esto es, que la siguiente afirmación es válida:

$$Suc(a) = Suc(\bar{a}) \Rightarrow a = \bar{a} \quad (4.6)$$

Para verificar esta afirmación se utilizará el método de contradicción. Con este fin, denote mediante  $b$  al valor común de  $Suc(a)$  y  $Suc(\bar{a})$ , esto es,

$$b = Suc(a) = Suc(\bar{a}),$$

y suponga que  $a \neq \bar{a}$ . En este caso, alguna de las dos alternativas ocurre:

$$a \prec \bar{a} \quad \text{o} \quad \bar{a} \prec a$$

y sin pérdida de generalidad alguna, se supondrá que la primera de estas posibilidades es válida. En este caso, se tiene que

$$a \prec \bar{a} \prec Suc(\bar{a}) = b,$$

donde la segunda relación de procedencia se debe a la definición de sucesor de  $\bar{a}$ . Sin embargo,  $b$  es también el sucesor de  $a$  de manera que la anterior relación desplegada implica que

$$\bar{a} \prec a \prec Suc(a) = b$$

y entonces  $\bar{a}$  sucede a  $a$ , pero es anterior a  $Suc(a) = b$ , lo cual se opone al hecho de que  $b$  es el sucesor de  $a$ . Esto muestra que el supuesto  $a \neq \bar{a}$  conduce a una contradicción, y por lo tanto la implicación (4.6) es válida. Para concluir, recuerde que el dominio de la función **Suc** es  $A \setminus \{M\}$ . Además,  $m$ , el elemento mínimo de  $A$  no puede ser el sucesor de ningún miembro de  $A$  pues, por la Definición 3.2(a),  $m$  no sucede a ningún elemento de  $A$ . Por lo tanto, **Suc** toma valores en  $A \setminus \{m\}$ , es decir, **Suc** transforma cada miembro de  $A \setminus \{M\}$  en un elemento de  $A \setminus \{m\}$ ;

Puesto que ambos conjuntos en este enunciado tienen el mismo número de elementos y la función **Suc** es uno a uno, se desprende que **Suc** transforma  $A \setminus \{M\}$  sobre  $A \setminus \{m\}$ , esto es, **Suc** es una biyección entre ambos conjuntos.

**Ejemplo 4.2** De nueva cuenta, sea  $A = \{A, *, 0, !\}$  dotado con el orden descrito en (4.4). Como ya se ha determinado con anterioridad, en este caso  $m = !$  es el elemento mínimo de  $A$ , mientras que  $M = 0$  es el miembro máximo. En este caso,

$$A \setminus \{M\} = \{A, *, !\} \quad \text{y} \quad A \setminus \{m\} = \{A, *, 0\}$$

Note que de acuerdo a (4.5),

$$Suc(!) = A$$

$$Suc(A) = *$$

$$Suc(*) = 0,$$

De manera que **Suc** establece una biyección entre  $\{A, *, !\}$  y  $\{A, *, 0\}$ , en concordancia con el Corolario 4.1.

## El Procedimiento General

Después de la teoría desarrollada en las secciones precedentes, ha llegado el momento de presentar el resultado central de este trabajo, a saber, un procedimiento que permite enumerar, o listar todos los elementos de un conjunto finito una sola vez. La premisa fundamental es que se ha definido un orden completo en el conjunto de interés, y la idea detrás del algoritmo es la siguiente: Inicie

la lista con un elemento mínimo del conjunto, y vaya agregando el sucesor del elemento que se ha anotado más recientemente, deteniendo el procedimiento cuando el último elemento incorporado no tenga sucesor, lo cual puede ocurrir sólo cuando dicho término es el máximo de los miembros del conjunto (vea el Teorema (4.1)).

**Algoritmo General de Listado**

**Datos:** (a) Un conjunto finito A con dos o más elementos, dotado con un orden completo.

(b) Se conoce el elemento mínimo de A, denotado por **m**, así como la función **Suc**, esto es, se puede determinar el sucesor de cualquier elemento que no sea el máximo, y además se puede detectar si un elemento dado no tiene sucesor, de tal forma que dicho elemento es el máximo de A.

**Fase 1:** [Inicialización.] Defina  $k = 1, a_1 = m$ , el elemento mínimo de A

**Fase 2:** [Fase de prueba.] Suponga que se tiene la lista (posiblemente) parcial de elemento  $a_1, a_2, \dots, a_k$  de A:

$$a_1, a_2, \dots, a_k$$

- (1) Si  $a_k$  tiene un sucesor, vaya a la Fase 3.
- (2) Si  $a_k$  no tiene sucesor, vaya a la Fase 4.

**Fase 3:** [Fase de iteración.](1) Incremente k en una unidad y defina  $a_{k+1}$  como el sucesor de  $a_k$ , esto es,

$$k \leftarrow k + 1, a_{k+1} = \text{Suc}(a_k)$$

(c) Retorne a la Fase 2.

**Fase 4:** [Fin de algoritmo.] La lista completa de los miembros de A es

$$a_1, a_2, \dots, a_k$$

Antes de demostrar la validez del algoritmo general de listado, se ilustrará su aplicación en el caso que ha sido utilizado anteriormente para ilustrar las ideas introducidas en este capítulo.

**Ejemplo 5.1.** Como antes, sea  $A = \{A, \emptyset, !\}$  dotado con la relación de orden (4.4). Como se verificó en el Ejemplo 3.2, el elemento mínimo de A es  $m = !$  y la función **Suc** está determinada en (4.5); vea el Ejemplo 4.5. La aplicación del algoritmo de listado es como sigue:

**Fase 1:**  $k = 1, a_1 = !$ , pues el elemento mínimo de A es  $!$

**Fase 2:** En este momento se tiene la lista parcial

!

Puesto que  $a_1 = !$  tiene sucesor, la ruta es hacia la Fase 3.

**Fase 3:** (1) Se incrementa k en 1 y se define  $a_2 = \text{Suc}(!) = A$ . Después de hacer estos cambios, se tiene.

$$k = 2, y a_1 = !, a_2 = A$$

**Fase 2:** En este momento  $k = 2$  y la lista parcial es  $!, A$

Puesto que  $a_2 = A$  tiene sucesor, la ruta es hacia la Fase 3.

**Fase 3:** (1) Se incrementa k en 1 y se define  $a_3 = \text{Suc}(A) = \emptyset$ . Después de hacer estos cambios, la situación se resume en

$$k = 3, y a_1 = !, a_2 = A, a_3 = \emptyset$$

(2) Se retorna a la fase 2.

**Fase 2:** Ahora el más reciente de los elementos incorporados a la lista es  $a_3 = \emptyset$ . Puesto que  $a_3$  no tiene sucesor, la ruta es hacia la Fase 4.

**Fase 3:** (1) Se incrementa k = 3 en 1 y se define  $a_4 = \text{Suc}(\emptyset) = 0$ . Después de hacer estos cambios, la situación se resume en.

$$k = 4, y a_1 = !, a_2 = A, a_3 = \emptyset, a_4 = 0$$

(2) Se retorna a la Fase 2.

**Fase 2:** El último elemento agregado a la lista es  $a_4 = 0$  y no tiene sucesor. Por lo tanto, la ruta es hacia la Fase 4.

**Fase 4:** El Algoritmo concluye: A tiene k = 4 elementos, y la lista de los elementos de A generada por el algoritmo es  $!, A, \emptyset, 0$ .

**Teorema 5.1.** Dado un conjunto finito A dotado con un orden completo, el algoritmo general de listado produce una enumeración de todos sus elementos.

**Demostración.** El argumento es por inducción en el número de elementos de A.

Se demostrará que:

Si A tiene m elementos, entonces el algoritmo termina después de m visitas a la Fase 2 (5.1)

Esta afirmación es claramente válida para  $m = 2$ . Suponga que (5.1) ocurre cuando un conjunto A dotado con un orden completo tiene  $m = n + 2$  elementos, y sea A un conjunto con  $n + 1$  miembros, entre los cuales se ha construido un orden total. Si  $M$  es el máximo de A defina el nuevo conjunto

$$\bar{A} = A \setminus \{M\},$$

el cual tiene  $n$  miembros (uno menos que el número de elementos de  $A$ ). Por lo tanto, aplicando el algoritmo al conjunto  $\bar{A}$  la hipótesis de inducción implica que, después de  $n$  visitas a la Fase 2, se habrá completado una lista de todos los miembros de  $\bar{A}$ . De hecho, en la visita número  $n$ , se encontrará que  $a$ , no tiene sucesor en  $\bar{A}$  y la bifurcación será a la Fase 4. Sin embargo, esto último ocurre si el interés se centra en  $\bar{A}$ . Si lo que se busca es listar los miembros del conjunto original  $A$ , se detectará que  $a$ , si tiene sucesor (el cual es  $M$ ) y la bifurcación será hacia la Fase 3, en el cual se pasará de  $k = n$  a  $k = n + 1$  y se definirá  $a' = \text{Suc}(a, 1) = M$ . Después de este paso, se retornará a la Fase 2 haciendo la visita  $n + 1$ , en la cual se detectará que  $M$  no tiene sucesor, y el flujo se dirigirá hacia la Fase 4, finalizando la lista de los miembros del conjunto original con  $n + 1$  pasos por la Fase 2. Este argumento verifica que (5.1) es válido con  $m = n + 1$ , y concluye el argumento de inducción.

### El Orden Lexicográfico

Cuando un conjunto finito  $A$  está dotado con un orden total (completo), el algoritmo propuesto en la sección anterior produce una lista de todos los miembros. De hecho, combinando el Teorema 5.1 con los comentarios plasmados en la sección 2, es claro que los problemas de generar una lista de los miembros de un conjunto finito y de definir una relación de orden total entre sus miembros, son problemas equivalentes. En las aplicaciones, la utilización del Teorema 5.1 para generar listados, depende de que se tenga a la mano una relación de orden completa. Los ejemplos presentados hasta ahora han tenido una finalidad ilustrativa, y han dejado a un lado la consideración de los conjuntos más complejos -la clase de permutaciones y la familia de subconjuntos- que son de interés primordial. Para dichos conjuntos, el problema de generar una lista de sus miembros tiene gran importancia en la *Estadística* no paramétrica; sin embargo, dichos conjuntos se ubican dentro de  $\mathbb{R}^p$  para algún entero  $p \geq 2$ , y es necesario introducir una relación de orden en este espacio multidimensional.

**Definición 6.1** La relación de orden lexicográfico en  $\mathbb{R}^p$  está determinada de la siguiente manera: Si  $a = (a_1, a_2, \dots, a_p)$  y  $b = (b_1, b_2, \dots, b_p)$  son los miembros arbitrarios de  $\mathbb{R}^p$ , entonces.

$a \prec b$  si y sólo si existe un entero  $k \leq p$  tal que

$$a_i = b_i, \text{ para } 1 \leq i < k, \text{ y } a_k < b_k$$

Verbalmente, esta relación puede describirse como sigue: Para comparar los miembros  $a, b \in \mathbb{R}^p$ , recorra  $a$  y  $b$  de izquierda a derecha simultáneamente, hasta encontrar la primera posición en que las componentes correspondientes de  $a$  y  $b$  difieran. En este caso, se declara que  $a \prec b$  si en dicha posición la componente de  $b$  es mayor que la de  $a$ , mientras que  $b \prec a$  en otro caso; por supuesto, si todas las componentes de  $a$  y de  $b$  coinciden, se tiene que  $a = b$ . Por ejemplo, considere el caso de  $p = 5$  y sean

$$a = (5, 5, 4, 7, 10), \quad b = (5, 5, 4, 10, 11).$$

En este caso, al recorrer  $a$  y  $b$  de izquierda a derecha, se observa que difieren en la posición número  $k = 4$ , y que la cuarta componente de  $b$  es 10, la cual es mayor a la cuarta componente de  $a$  (la cual es 7):

$$a_i = b_i, \quad 1 \leq i < 4, \quad a_4 < b_4$$

De acuerdo a la Definición 6.1, se desprende que  $a \prec b$ . El siguiente teorema establece que la relación introducida en la Definición 6.1 es, efectivamente, una relación de orden total.

**Teorema 6.1.** La ordenación lexicográfica en  $\mathbb{R}^p$  es completa (total) en el sentido de la Definición 3.1

**Demostración.** Se verificará que el orden lexicográfico tiene las propiedades de transitividad y completez:

(a) [Transitividad.] Suponga que  $a, b, c \in \mathbb{R}^p$  son tales que

$$a \prec b, \text{ y } b \prec c \tag{6.1}$$

En tal caso, existen enteros  $k_1$  y  $k_2$  entre 1 y  $p$  tales que

$$a_i = b_i, \quad 1 \leq i < k_1, \quad a_{k_1} < b_{k_1}, \quad \text{y} \quad b_i = c_i, \quad 1 \leq i < k_2, \quad b_{k_2} < c_{k_2} \tag{6.2}$$

Considere ahora los siguientes casos: (1)  $k = k_2 = k_1$ . En estas circunstancias, (6.2) implica que  $a_i = c_i$  siempre que  $1 \leq i < k$  mientras que  $a_k = c_k < b_k = b_{k_2} < c_{k_2} = c_k$ . esto es, es la primera posición para la cual los componentes de **a** y **c** difieren, la componente **c** es mayor. Por lo tanto,

$$a \prec c.$$

(2)  $k_1 < k_2$ . En este caso, defina  $k = k_2$ . Utilizando (6.2), se tiene que

$$a_i = b_i, \quad 1 \leq i < k, \quad y \quad a_k < b_k$$

Además, debido a que  $k < k_1$ , se tiene que  $b_i = c_i$ ,  $1 \leq i \leq k$ , por lo tanto, la última relación desplegada implica que

$$a_i = c_i, \quad 1 \leq i < k, \quad y \quad a_k < c_k$$

de tal forma que  $a \prec c$ . (3)  $k_1 > k_2$ . En este caso, defina  $k = k_2$ . De nueva cuenta, utilizando (6.2), se obtiene, puesto que  $k_1 > k = k_2$ , que

$$a_i = b_i, \quad 1 \leq i \leq k,$$

mientras que

$$a_i = c_i, \quad 1 \leq i < k, \quad b_k < c_k.$$

Por lo tanto, a partir de estas relaciones desplegadas se obtiene

$$a_i = b_i, \quad 1 \leq i < k, \quad a_k = b_k < c_k,$$

de tal forma que,  $a \prec c$ . Este argumento ha demostrado que, en cualquier circunstancia, se tiene que  $a \prec c$  siempre que (6.1) sea válida, esto es,

$$a \prec b, \quad y \quad b \prec c \Rightarrow a \prec c,$$

estableciendo la propiedad de transitividad.

(b) [Completez.] Dados dos vectores  $a, b \in \mathbb{R}^n$ , debe demostrarse que exactamente una de las siguientes alternativas ocurre:

$$a \prec b, \quad a = b, \quad o \quad b \prec a \quad (6.3)$$

Suponga que  $a \neq b$ . En este caso, **a** y **b** difieren en alguna componente; defina  $k$  como el primer entero en el que las componentes correspondientes de **a** y **b** son distintas, esto es,

$$a_i = b_i, \quad 1 \leq i < k, \quad a_k \neq b_k.$$

Si  $a_k < b_k$ , la Definición 6.1 implica  $a \prec b$ , mientras que si  $a_k > b_k$ , entonces  $b \prec a$ . Por lo tanto, se ha demostrado que si  $a \neq b$ , entonces exactamente una de las otras dos relaciones en (6.3) ocurre; esto establece que el orden lexicográfico es completo, y concluye la demostración del Teorema.

## Conclusión

En este trabajo se ha estudiado el problema general de listado (enumeración) para un conjunto finito **A**. La presentación puso de manifiesto la equivalencia entre este problema y la construcción de una relación de orden total entre los miembros de **A**, y se formuló un algoritmo general de listado, el cual permite enumerar todos los miembros de un conjunto finito arbitrario, tan pronto como se haya definido una relación de orden total entre sus miembros. Sin embargo, debe señalarse que la aplicación efectiva de dicho procedimiento depende de las siguientes condiciones:

- (a) Construir un orden total en el conjunto de interés;
- (b) Determinar el elemento mínimo del conjunto;
- (c) Especificar completamente la función **Suc**, la cual transforma un miembro de **A** que no sea el máximo, en su sucesor.

El punto (a) fue considerado en la Sección 6, donde se introdujo el orden lexicográfico en  $\mathbb{R}^n$ , y se demostró que es un orden total, de manera que un conjunto contenido en  $\mathbb{R}^n$ , siempre puede considerarse dotado del orden lexicográfico. Sin embargo, los aspectos (b) y (c) son específicos del conjunto **A** bajo consideración, esto es, la dificultad de determinar el mínimo de **A** y la función **Suc** dependen de las características de cada conjunto contenido en  $\mathbb{R}^n$ , e implican un análisis detallado cuya complejidad depende del caso específico analizado.

## Literatura citada

- 1 Dolciani, M. P., E. Beckenbach. E. Reilich (1986), Algebra II Publicaciones culturales, 1976, México, D.F.
- 2 Munkres, J. R. (1975), Topology, A First Course, Prentice – Hay, Englewood Cliffs, New Jersey.
- 3 Papadopoulos, J. S. (1997), Calculate PI, ABCCode, 2, June, 1997.
- 4 Suppes, P. (1971), Set Theory, Van – Nostrand, New York.



# Caracterización y Aplicación de la Distribución Birnbaum-Saunders como Modelo de Tiempos de Vida

Edgar Espinosa Cuello

Mario Cantú Sifuentes

Departamento de Estadística y Cálculo, Universidad Autónoma Agraria Antonio Narro, México

Víctor Leiva Sánchez

Departamento de Estadística, Universidad de Valparaíso, Chile.

**Abstract.** *Characterization and Application of the Birnbaum-Saunders Distribution as Life Time Model.* The Birnbaum-Saunders distribution was developed originally to modeling lifetimes of materials exposed to fatigue tests. In this work three examples with real data, where such a distribution is applied, are studied. It is taken into account both cases: complete and censored data. As a fourth example, a computational algorithm for simulation of data from a Birnbaum-Saunders distribution is presented.

**Key words:** Birnbaum-Saunders distribution, Fatigue of materials, Goodness-of-fit tests.

**Resumen.** La distribución Birnbaum-Saunders fue desarrollada originalmente para modelar tiempos de vida de materiales expuestos a pruebas de fatiga. En el presente trabajo se estudian tres ejemplos con datos reales donde se aplica dicha distribución. Se toman en cuenta tanto el caso de datos completos como el caso de datos censurados. En un cuarto ejemplo se presenta un algoritmo computacional para generar datos provenientes de una distribución Birnbaum-Saunders.

**Palabras clave:** Distribución Birnbaum-Saunders, Fatiga de materiales, Pruebas de bondad de ajuste.

## Introducción

En 1969, motivados por los problemas de vibración encontrados en los nuevos aviones comerciales y por los problemas de fatiga de materiales, Birnbaum y Saunders (1969a) presentaron un modelo probabilístico ingenioso para tiempos de vida de estructuras bajo carga dinámica. La distribución Birnbaum-Saunders (BS) fue derivada a partir de un modelo que muestra que las fallas se deben al desarrollo y crecimiento de una fractura dominante. Así, la distribución BS describe totalmente el tiempo de falla que ha transcurrido cuando cierta clase de daño acumulado excede un umbral. Su función de densidad de probabilidad (FDP) es de la forma

$$f(t) = \frac{1}{\alpha\beta} \xi'(t/\beta) \phi[\alpha^{-1} \xi(t/\beta)],$$

donde  $\xi(t) = t^a - t^{-1/2}$ ,  $\xi'(t) = d\xi(t)/dt$ ,  $t, a, \beta > 0$  y  $\phi$  es la FDP de la distribución Normal estándar. En palabras, una variable aleatoria (v.a.) positiva  $T$  se dice que se distribuye Birnbaum-Saunders, denotado

por  $T \sim BS(\alpha, \beta)$  si la transformación de la v.a.  $Z = \alpha^{-1}[(T/\beta)^{1/2} - (T/\beta)^{-1/2}]$  tiene una distribución Normal estándar. El parámetro  $\beta$  es un parámetro de escala y también la mediana de la distribución. El parámetro  $\alpha$  es un parámetro de forma para la distribución.

Algunas propiedades básicas de la distribución BS son:

i  $E(T) = \beta \left( 1 + \frac{\alpha^2}{2} \right),$

ii  $VAR(T) = (\alpha\beta)^2 \left( 1 + \frac{5\alpha^2}{4} \right),$

iii si  $T \sim BS(\alpha, \beta)$ , entonces  $cT \sim BS(\alpha, \beta^c)$

y  $cT \sim BS(\alpha, c\beta)$  para cualquier real  $c > 0$ ,

v si  $Z = \alpha^{-1} \left[ \left( \frac{T}{\beta} \right)^{1/2} - \left( \frac{T}{\beta} \right)^{-1/2} \right]$  entonces

$$T = \beta \left\{ \frac{\alpha Z}{2} + \left[ \left( \frac{\alpha Z}{2} \right)^2 + 1 \right]^{1/2} \right\}^2,$$

donde  $E(T)$  y  $VAR(T)$  denotan la media y la varianza, respectivamente.

Birnbaum y Saunders (1969b) obtuvieron los estimadores de máxima verosimilitud para  $a$  y  $\beta$ . Engelhardt et al. (1981) desarrollaron intervalos de confianza y pruebas de hipótesis usando muestras completas grandes y pequeñas. Los procedimientos para muestras grandes están basados en la normalidad asintótica de los estimadores de verosimilitud máxima, mientras que en el caso de las muestras pequeñas los métodos se basan en simulación. Desmond (1983, 1985 y 1986) obtuvo una derivación de la distribución BS más general y estableció una conexión entre la distribución Inversa Gaussiana y la distribución BS. Rieck y Nedelman (1991) propusieron un modelo log-lineal para la distribución BS y desarrollaron métodos de estimación. McCarter (1999) discute la estimación de parámetros para el caso de muestras censuradas tipo II. Hay varias generalizaciones de la distribución BS, entre ellas están la propuesta por Owen y Padgett (1999) quienes hacen una generalización de dos a tres parámetros junto con su caracterización y algunos ejemplos. Díaz y Leiva (2002) presentan otra generalización de la distribución BS, generándola desde una distribución de Contornos Elípticos, la caracterizan y obtienen una familia de distribuciones de vida en situaciones diferentes y variadas, tales como distribuciones de vida que no tienen momentos, cuando las colas son más pesadas o menos pesadas, o cuando la distribución sea bimodal.

El objetivo de este trabajo es aplicar métodos estadísticos relacionados con la distribución BS a problemas de fatiga en materiales. Se darán cuatro ejemplos, tres con datos reales tomados de la literatura existente relacionada con el tema y un cuarto en donde se presenta un ejemplo basado en simulación.

### Ejemplo 1

Para este ejemplo se usan los datos que emplearon originalmente Birnbaum y Saunders para sustentar la distribución que lleva su nombre. Se presentan tres conjuntos de datos que corresponden a tiempos de vida en ciclos  $\times 10^{-3}$  de piezas de aluminio 6061-T6 cortadas en un ángulo paralelo a la dirección de rotación y oscilando a 18 ciclos por segundo a una presión máxima de 31000, 26000 y 21000psi, donde psi (Pounds per

Square Inch) es una medida de presión. Estos conjuntos de datos son completos (y los denotaremos por T) y se presentan en los Cuadros 1, 2 y 3, para 31000, 26000 y 21000psi, con tamaños de muestras de 101, 102 y 101, respectivamente.

**Cuadro 1.** Tiempos de vida en ciclos  $\times 10^{-3}$  de piezas de aluminio 6061-T6 sujetas a una presión máxima de 31000 psi.

70	90	96	97	99	100	103	104	104	105
107	108	108	108	109	109	112	112	113	114
114	114	116	119	120	120	120	121	121	123
124	124	124	124	124	128	128	129	129	130
130	130	131	131	131	131	131	132	132	132
133	134	134	134	134	134	136	136	137	138
138	138	139	139	141	141	142	142	142	142
142	142	144	144	145	146	148	148	149	151
151	152	155	156	157	157	157	157	158	159
162	163	163	164	166	166	168	170	174	196
212									

**Cuadro 2.** Tiempos de vida en ciclos  $\times 10^{-3}$  de piezas de aluminio 6061-T6 sujetas a una presión máxima de 26000 psi.

233	258	268	276	290	310	312	315	318	321
321	329	335	336	338	338	342	342	342	344
349	350	350	351	351	352	352	356	358	358
360	362	363	366	367	370	370	372	372	374
375	376	379	379	380	382	389	389	395	396
400	400	400	403	404	406	408	408	410	412
414	416	416	416	420	422	423	426	428	432
432	433	433	437	438	439	439	443	445	445
452	456	456	460	464	466	468	470	470	473
474	476	476	486	488	489	490	491	503	517
540	560								

**Cuadro 3.** Tiempos de vida en ciclos  $\times 10^{-3}$  de piezas de aluminio 6061-T6 sujetas a una presión máxima de 21000 psi

370	706	716	746	785	797	844	855	858	886
886	930	960	988	999	1000	1010	1016	1018	1020
1055	1085	1102	1102	1108	1115	1120	1134	1140	1199
1200	1200	1203	1222	1235	1238	1252	1258	1262	1269
1270	1290	1293	1300	1310	1313	1315	1330	1355	1390
1416	1419	1420	1420	1450	1452	1475	1478	1481	1485
1502	1505	1513	1522	1552	1530	1540	1560	1567	1578
1594	1602	1604	1608	1630	1642	1674	1730	1750	1750
1763	1768	1781	1782	1792	1820	1868	1888	1890	1893
1895	1910	1923	1924	1945	2023	2100	2130	2215	2268
2440									

### Análisis Exploratorio

Para la variable T, se procedió a aplicar tres criterios de bondad de ajuste.

#### Gráficos P-P

Si se tienen n observaciones que se cree que provengan de una distribución F, se puede construir un gráfico de las probabilidades empíricas versus probabilidades teóricas de dichas observaciones. Si estos puntos están aproximadamente alineados será un indicio de que las observaciones provienen de dicha distribución.

A continuación se presentan los gráficos P-P para la v.a. T con una presión máxima por ciclo de: 31000psi (Figura 1), 26000psi (Figura 2) y 21000psi (Figura 3) para las distribuciones de vida usadas comúnmente para

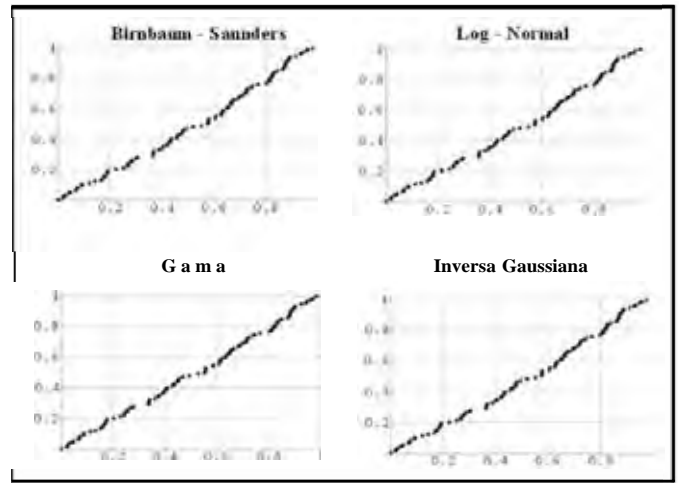


Figura 3. Gráficas P-P de tiempos de vida de piezas de aluminio 6061-T6 sujetas a una presión máxima de 21000 psi para las distribuciones Birnbaum-Saunders, Lognormal, Gamma e Inversa Gaussiana.

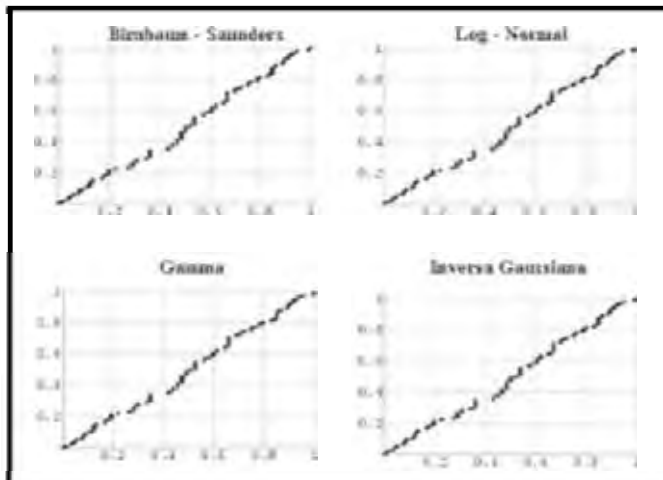


Figura 1. Gráficas P-P de tiempos de vida de piezas de aluminio 6061-T6 sujetas a una presión máxima de 31000 psi para las distribuciones Birnbaum-Saunders, Lognormal, Gamma e Inversa Gaussiana

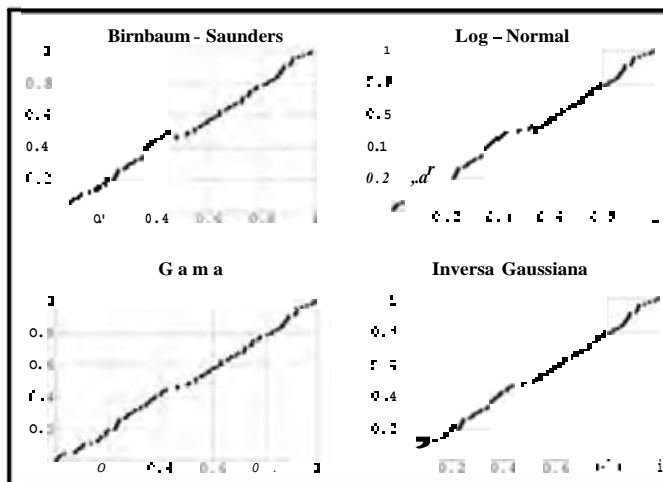


Figura 2. Gráficas P-P de tiempos de vida de piezas de aluminio 6061-T6 sujetas a una presión máxima de 26000psi para las distribuciones Birnbaum-Saunders, Log-normal, Gamma e Inversa Gaussiana.

modelar fallas debido a fatiga: Birnbaum-Saunders, Lognormal, Gamma e Inversa Gaussiana.

Según este criterio, y para todos los niveles máximos de presión indicados, no hay evidencia que permita descartar alguna de las distribuciones de vida por fatiga que se utilizan frecuentemente. Para tomar una determinación se utilizarán otros criterios.

#### Prueba de Kolmogorov-Smirnov

La prueba de Kolmogorov-Smirnov (KS) se utiliza para decidir si una muestra proviene o no de una población con una distribución específica  $F(x)$ , basándose en la función de distribución empírica (FDE). Específicamente, dados n puntos  $x_1, x_2, \dots, x_n$ , se define la FDE  $F_n(x) = m(t)/n$ , donde  $m(t)$  es el número de puntos menores que  $x$ , y es una función que aumenta con n.

La prueba KS se basa en la distancia máxima entre  $F(x)$  y  $F_n(x)$ . Más concretamente, ésta consiste en contrastar

$$H_0 : F(x) = F_n(x) \text{ versus } H_1 : F(x) \neq F_n(x),$$

donde  $F_n(x)$  es la distribución que se sospecha siguen los datos.

La estadística de prueba viene dada por

$$KS = \max_x |F_n(x) - F(x)|$$

La hipótesis con respecto a la forma distribucional de los datos se rechaza si la estadística de prueba KS es mayor que el percentil  $1-\alpha$  de la distribución de Kolmogorov-Smirnov, la cual viene implementada en la mayoría de los software estadísticos, por ejemplo S-Plus.

A continuación se presentan las pruebas KS para la v.a. T con una presión máxima por ciclo de: 31000psi

(Cuadro 4), 26000psi (Cuadro 5) y 2 1000psi (Cuadro 6) para las distribuciones de vida por fatiga Birnbaum-Saunders, Lognormal, Gamma e Inversa Gaussiana, donde el criterio es tomar la distribución que tenga el mayor valor p.

**Cuadro 4.** Concentrado de la prueba de bondad de ajuste de Kolmogorov-Smirnov (KS) para los datos de tiempos de vida de piezas de aluminio 6061-T6 sujetas a una presión máxima de 31000 psi para las distribuciones Birnbaum-Saunders, Lognormal, Gamma e Inversa Gaussiana

Distribución	Estadístico KS	p-value
Gamma	0.0727	0.659
Lognormal	0.0842	0.075
Birnbaum-Saunders	0.1095	0.000
Inversa Gaussiana	0.3849	0.000

**Cuadro 5.** Concentrado de la prueba de bondad de ajuste de Kolmogorov-Smirnov (KS) para los datos de tiempos de vida de piezas de aluminio 6061-T6 sujetas a una presión máxima de 26000 psi para las distribuciones Birnbaum-Saunders, Lognormal, Gamma e Inversa Gaussiana

Distribución	Estadístico KS	p-value
Inversa Gaussiana	0.0369	1.000
Birnbaum-Saunders	0.0370	1.000
Gamma	0.0442	0.989
Lognormal	0.0539	0.500

**Cuadro 6.** Concentrado de la prueba de bondad de ajuste de Kolmogorov-Smirnov (KS) para los datos de tiempos de vida de piezas de aluminio 6061-T6 sujetas a una presión máxima de 21000 psi para las distribuciones Birnbaum-Saunders, Lognormal, Gamma e Inversa Gaussiana

Distribución	Estadístico KS	p-value
Gamma	0.0581	0.885
Birnbaum-Saunders	0.0750	0.543
Lognormal	0.0742	0.500
Inversa Gaussiana	0.3655	0.330

Según este criterio, sólo en el segundo conjunto de datos (Cuadro 5) la distribución Birnbaum-Saunders fue la distribución de vida por fatiga que mejor se ajusta a los datos.

### Criterio de Información de Schwartz

Este criterio, para un modelo con parámetro  $\theta \in \mathbb{R}^p$ , está dado por:

$$SIC = -2\text{Log}[f(y|\theta)] + p\text{log}(n),$$

donde  $f(y|\theta)$  es la densidad asociada a los datos y n es la cantidad de observaciones. Así, cuando se disponga de varios modelos hipotéticos que se puedan ajustar a un conjunto de datos, un criterio es seleccionar aquel modelo que tenga el menor SIC.

A continuación se utiliza este criterio para los tiempos de vida T con estrés máximo por ciclo de: 31000psi (Cuadro 7), 26000psi (Cuadro 8) y 21000psi (Cuadro 9) para las distribuciones de vida por fatiga Birnbaum-Saunders, Lognormal, Gamma e Inversa Gaussiana.

**Cuadro 7.** Resultados del criterio de información de Schwartz de selección de modelos para los datos de tiempos de vida de piezas de aluminio 6061-T6 sujetas a una presión máxima de 31000 psi para las distribuciones Birnbaum-Saunders, Lognormal, Gamma e Inversa Gaussiana

Distribución	SIC
Gamma	921.886
Lognormal	923.468
Birnbaum-Saunders	923.771
Inversa Gaussiana	923.802

**Cuadro 8.** Resultados del criterio de información de Schwartz de selección de modelos para los datos de tiempos de vida de piezas de aluminio 6061-T6 sujetas a una presión máxima de 26000 psi para las distribuciones Birnbaum-Saunders, Lognormal, Gamma e Inversa Gaussiana

Distribución	SIC
Gamma	1142.39
Lognormal	1144.56
Birnbaum-Saunders	1144.65
Inversa Gaussiana	1144.68

**Cuadro 9.** Resultados del criterio de información de Schwartz de selección de modelos para los datos de tiempos de vida de piezas de aluminio 6061-T6 sujetas a una presión máxima de 21000 psi para las distribuciones Birnbaum-Saunders, Lognormal, Gamma e Inversa Gaussiana

Distribución	SIC
Gamma	1503.73
Lognormal	1510.42
Birnbaum-Saunders	1511.97
Inversa Gaussiana	1512.36

Según este criterio, en ninguno de los casos la distribución BS fue la distribución de vida por fatiga que mejor se ajusta a los datos.

Como conclusión de este ejemplo, podemos decir que los datos que emplearon Birnbaum y Saunders para justificar la distribución que lleva su nombre, se ajustan mejor con otros modelos. En la mayoría de los casos, parece ser la distribución Gamma la que mejor ajusta los datos. Sin embargo, debido a su génesis, la distribución BS se utiliza con más frecuencia para modelar tiempos de vida debido a fatiga.

### Ejemplo 2

Considere los datos de fatiga biaxial reportados por Brown y Miller (1978) y el modelo log-lineal  $Y = X + \epsilon$ , propuesto por Rieck y Nedelman (1991). En el ejemplo de Brown y Miller la respuesta, N, es el número de ciclos hasta que ocurre la falla y la variable independiente, W, es el esfuerzo por ciclo (medida en MJ/m<sup>3</sup>). Se consideraron 46 observaciones para establecer un modelo adecuado de predicción, las que se presentan en el Cuadro 10.

**Cuadro 10.** Datos de tiempo a fatiga de Brown y Miller (1978).

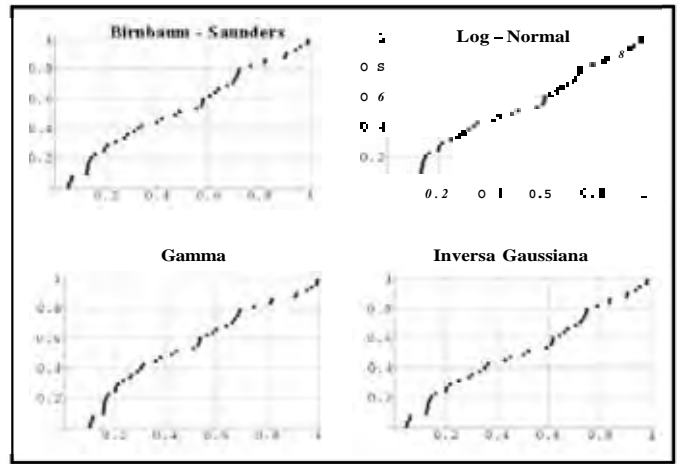
W	N	W	N	W	N	W	N
11.5	3280	24.0	804	40.1	750	60.3	283
13.0	5046	24.6	1093	40.1	316	60.5	212
14.3	1563	25.2	1125	43.0	456	62.1	327
15.6	4707	25.5	884	44.1	552	62.8	373
16.0	977	26.3	1300	46.5	355	66.5	125
17.3	2834	27.9	852	47.3	242	67.0	187
19.3	2266	28.3	580	48.7	190	67.1	135
21.1	2208	28.4	1066	52.9	127	67.9	245
21.5	1040	28.6	1114	56.6	185	68.8	137
22.6	700	30.9	386	59.9	255	75.4	200
22.6	1583	31.9	745	60.2	195	100.5	190
24.0	482	34.5	736				

### Análisis Exploratorio

Para la variable N, número de ciclos hasta que ocurre la falla, se aplicaron los tres criterios de bondad de ajuste usados en el ejemplo 1. Los resultados fueron los siguientes.

#### Gráficos P-P

A continuación se presentan los gráficos P-P para los datos de fatiga biaxial de Brown y Miller (1978) para las



**Figura 4.** Gráficas P-P de los tiempos de vida de Brown y Miller (1978) para las distribuciones Birnbaum-Saunders, Lognormal, Gamma e Inversa Gaussiana.

distribuciones de vida por fatiga Birnbaum-Saunders, Lognormal, Gamma e Inversa Gaussiana.

Según este criterio, aunque casi todas las distribuciones de vida por fatiga que se presentan se ajustan bien al conjunto de datos, la distribución BS es una de las que mejor se comporta.

### Prueba de Kolmogorov-Smirnov

A continuación se presenta la prueba KS para los datos de fatiga biaxial de Brown y Miller (1978) para las distribuciones de vida por fatiga Birnbaum-Saunders, Lognormal, Gamma e Inversa Gaussiana.

**Cuadro 11.** Concentrado de la prueba de bondad de ajuste de Kolmogorov-Smirnov (KS) para los datos de Brown y Miller (1978) para las distribuciones Birnbaum-Saunders, Lognormal, Gamma e Inversa Gaussiana

Distribución	Estadístico KS	p-value
Birnbaum-Saunders	0.0228	0.542
Inversa Gaussiana	0.0499	0.526
Lognormal	0.0917	0.500
Gamma	0.8953	0.000

Según este criterio, la distribución de vida por fatiga que mejor se ajusta a los datos es la distribución BS.

### Criterio de Información de Schwartz

A continuación se presenta el criterio de información de Schwartz para los datos de fatiga biaxial de Brown y Miller (1978) para las distribuciones de vida por fatiga Birnbaum-Saunders, Lognormal, Gamma e Inversa Gaussiana.

**Cuadro 12.** Resultados del criterio de información de Schwartz de selección de modelos para los datos de Brown y Miller (1978) para las distribuciones Birnbaum-Saunders, Lognormal, Gamma e Inversa Gaussiana

Distribución	SIC
Inversa Gaussiana	717.391
Birnbaum-Saunders	718.530
Lognormal	720.538
Gamma	729.490

Con este criterio, la distribución de vida por fatiga que mejor se ajusta a los datos es la distribución Inversa Gaussiana, seguida muy de cerca por la distribución BS.

En base a los tres criterios empleados, se concluye que la distribución que mejor se ajusta a los datos de fatiga biaxial es la distribución BS, cuyas estimaciones de sus parámetros son  $\hat{\alpha} = 1.08574$  y  $\hat{\beta} = 600.91$ .

### Análisis Confirmatorio

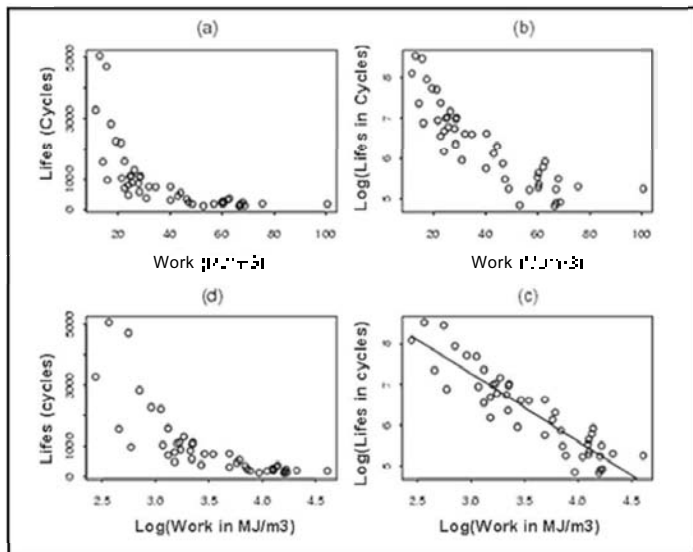
En la Figura 5(a), 5(b), 5(c) y 5(d) se presentan los diagramas de dispersión de W versus N, W versus Log(N), Log(W) versus N y Log(W) versus Log(N), respectivamente.

Se ve que el modelo exponencial (Figura 5(d)) es el más adecuado, esto es, el modelo log-lineal

$$\log(N_i) = \beta_1 + \beta_2 \log(W_i) + \epsilon_i, \quad i = 1, 2, \dots, 46$$

donde  $\epsilon_i \sim N(\alpha, \sigma^2)$ . Las estimaciones de verosimilitud máxima de  $\beta = \{\beta_1, \beta_2, \alpha\}$  son (con sus errores estándar estimados entre paréntesis):

$$\hat{\beta}_1 = 12.289 (0.406) \quad y \quad \hat{\beta}_2 = -1.673 (0.113)$$



**Figura 5.** Para los datos de Brown y Miller (1978) se presentan los diagramas de dispersión; en la Fig. 5 (a) el de W v.s N; en la 5(b) el de W v.s Log(N); en la 5(c) el de Log(W) v.s. N y en la 5 (d) el de Log(w) v.s Log(N).

ambos coeficientes  $\{\hat{\beta}_1, \hat{\beta}_2\}$  son estadísticamente significativos, con valores-p de  $4,17416 \times 10^{-31}$  y  $9.92861 \times 10^{-19}$ , respectivamente. El valor-p de la ANOVA es  $9,92861 \times 10^{-19}$ , el cual también es altamente significativo. El porcentaje de variabilidad explicado por el modelo es del 83.3%.

### Ejemplo 3

En el siguiente ejemplo se consideran los datos de fatiga biaxial reportados por Brown y Miller en 1978, en los cuales, para aplicar la metodología de datos censurados, se consideró una censura tipo II al 20%, obteniéndose los siguientes datos.

**Cuadro 13.** Datos de fatiga biaxial de Brown y Miller, con censura tipo II al 20%. \* indica censura por la derecha.

125	127	135	137	185	187	190	190
195	200	212	242	245	255	283	316
327	355	373	386	456	482	552	580
700	736	745	750	804	852	884	977
1040	1066	1093	1114	1125	1125*	1125*	1125*
1125*	1125*	1125*	1125*	1125*	1125*	1125*	1125*

### Análisis Estadístico

En el Cuadro 14 se presentan las estimaciones de  $\alpha$  y  $\beta$  de la distribución BS, a partir de una muestra censurada tipo II al 20%, junto con sus intervalos de confianza al 95%. En el Cuadro 15, con carácter referencial, se presentan los mismos resultados para los datos completos de Brown y Miller.

**Cuadro 14.** Estimación de los parámetros y sus intervalos de confianza al 95% de la distribución BS para los datos de fatiga biaxial de Brown y Miller con censura tipo II al 20 %.

Parámetro	Estimación	Límite inferior	Límite superior
$\alpha$	0.9516	0.9137	0.9894
$\beta$	538.0290	516.0131	560.0448

**Cuadro 15.** Estimación de los parámetros y sus intervalos de confianza al 95% de la distribución BS para los datos de fatiga biaxial de Brown y Miller completos.

Parámetro	Estimación	Límite inferior	Límite superior
$\alpha$	1.0857	0.0000	1.1184
$\beta$	600.9100	593.3405	608.4794

### Ejemplo 4

#### Procedimiento de Generación de Números Aleatorios BS

Sea  $T$  una variable aleatoria con distribución BS con parámetros de forma  $\alpha$  y de escala  $\beta$ . Denote su función de distribución por

$$F_T(t; \alpha, \beta) = \Phi(g(t; \alpha, \beta)),$$

donde

$$g(t, \alpha, \beta) = \frac{1}{\alpha} \left( \sqrt{\frac{t}{\beta}} - \sqrt{t} \right)$$

y  $\Phi$  es la función de distribución Normal estándar. Esto implica que la variable aleatoria  $g(t, \alpha, \beta)$  sigue distribución Normal estándar. Como se sabe,  $g(t, \alpha, \beta)$  es estrictamente creciente y por lo tanto tiene inversa, dada por  $g^{-1}(t; \alpha, \beta)$ . Sea  $Z$  una variable aleatoria que sigue una distribución Normal estándar. Entonces,

$$g(t; \alpha, \beta) \xrightarrow{u} Z \Rightarrow T \xrightarrow{u} g^{-1}(Z; \alpha, \beta) = \beta \left( 1 + \frac{\alpha^2 Z^2}{2} + \alpha Z \sqrt{1 + \frac{\alpha^2 Z^2}{4}} \right)$$

donde  $T$  fue dada anteriormente.

Ahora, si se dispone de un generador de números aleatorios con distribución Normal estándar, entonces se puede considerar el siguiente procedimiento:

- i. Generar una muestra aleatoria de tamaño  $n$  con distribución Normal estándar, la cual se denotará por  $Z_1, Z_2, \dots, Z_n$ .
- ii. Tome

$$T_i = \beta \left( 1 + \frac{\alpha^2 Z_i^2}{2} + \alpha Z_i \sqrt{1 + \frac{\alpha^2 Z_i^2}{4}} \right), i = 1, 2, \dots, n$$

De este modo,  $T_1, T_2, \dots, T_n$  es una muestra aleatoria de tamaño  $n$  proveniente de una distribución BS con parámetros  $\alpha$  y  $\beta$ .

#### Datos Simulados

Con el procedimiento visto en la sección anterior, se generaron tres conjuntos de 10, 100 y 1000 números aleatorios tomando como parámetro  $\alpha = 0,170385$  y  $\beta = 131.819$ . Como muestra, se presentan en el Cuadro 14 un conjunto de 100 datos generados con el procedimiento anterior ordenados ascendentemente.

En el Cuadro 15 se presentan los verdaderos valores de los parámetros  $\alpha$  y  $\beta$  (valores que se consideraron para el procedimiento de generación de números aleatorios) comparados con los valores estimados desde las muestras de tamaños 10, 100 y 1000. En el Cuadro 16 se muestra el error relativo entre los verdaderos valores y sus estimaciones.

Las gráficas de la función de distribución BS con los parámetros utilizados para la generación de números aleatorios se presenta en la Figura 6, donde los puntos negros representan la frecuencia acumulada para los datos generados aleatoriamente.

**Cuadro 16.** Cien datos aleatorios provenientes de una distribución BS.

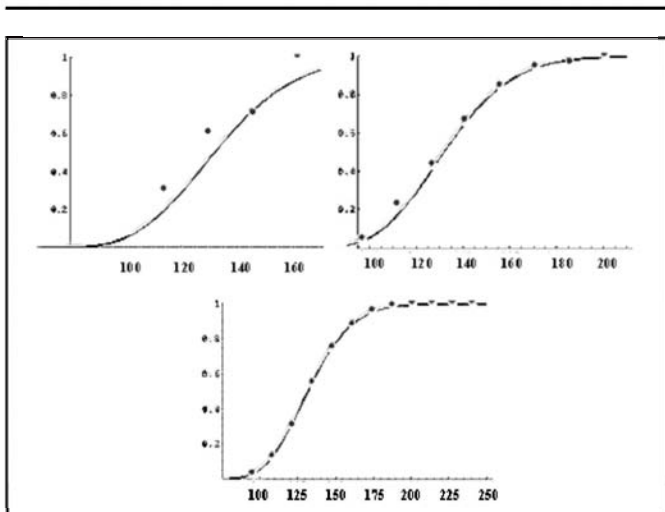
82.60	88.78	92.59	96.40	97.42	100.72	100.94	101.10	103.13
104.83	104.99	105.97	106.23	106.76	107.60	107.85	109.35	110.45
110.66	111.34	112.05	112.13	112.49	113.98	114.43	116.31	116.82
116.95	116.96	117.23	117.37	117.42	118.23	118.52	119.40	121.21
121.37	123.05	123.20	124.18	124.22	125.02	125.74	127.13	127.85
129.49	129.69	130.05	130.07	131.16	131.33	132.13	132.50	132.65
132.95	133.75	136.37	136.76	136.79	137.04	137.18	137.76	138.45
138.67	139.79	141.34	142.30	144.46	145.18	147.87	148.32	149.11
149.50	149.97	150.46	150.67	151.45	151.72	152.55	152.86	153.21
153.58	155.02	156.17	157.78	158.13	158.44	159.78	165.28	166.21
167.84	170.07	170.90	170.98	174.12	174.91	187.15	189.55	198.94
201.07								

**Cuadro 17.** Valores reales y estimados de los parámetros de la distribución BS para muestras simuladas de tamaño 10, 100 y 1000.

Tamaño muestral	Valor real		Valor estimado	
	$a$	$\beta$	$a$	$\beta$
10	0.170385	131.819	0.163712	126.881
100	0.170385	131.819	0.182288	131.170
1000	0.170385	131.819	0.167278	132.971

**Cuadro 18.** Error relativo de los estimadores de los parámetros del modelo BS usando muestras aleatorias simuladas de tal distribución para tamaños de muestra 10, 100 y 1000.

Tamaño	Para $a$	Para $\beta$
10	0.0391642	0.0374605
100	0.0698594	0.00492342
1000	0.0182352	0.00873926



**Figura 6.** Comparación entre la distribución acumulada empírica (puntos en negro) y la teórica (línea continua) para los datos simulados de la distribución BS con tamaños de muestra 10, 100

## Conclusiones

En el trabajo se utilizaron métodos para analizar datos desde una perspectiva de datos completos (ej. 1 y 2) y censurados (ej. 3). Para desarrollar las aplicaciones se implementaron programas en los software S-plus y Matemática.

En el ejemplo uno se descubrió que los datos de fatiga que se utilizaron en el artículo original de Birnbaum y Saunders de 1969 se ajustan mejor a otras distribuciones, lo que se confirmó a través de tres métodos de bondad de ajuste.

Es importante destacar que cuando el problema de estimación de la distribución BS generalizada planteada por Díaz y Leiva (2002) esté resuelto, sería interesante realizar nuevamente la bondad de ajuste y observar si alguna de las distribuciones de esa extensa familia de distribuciones de vida pudiera ajustarse mejor que las que existen hasta ahora.

## Literatura Citada

- Birnbaum, Z.W., Saunders, S.C. (1969a). A new family of life distributions. Journal of applied probability 6: 319-327.
- (1969b). Estimation for a family of life distribution with applications to fatigue. Journal of applied probability 6: 328-347.
- Brown, M.W., Miller, K.J. (1978). Biaxial fatigue data. Report CEMR1178, University of Sheffield. Department of mechanical engineering.
- Desmond, A.F. (1983). A local maxim of stationary stochastic processes and stochastic modeling of fatigue. Ph.D. Thesis, University of Waterloo.
- (1985). Stochastic models of failure in random environments. Canadian journal of statistics 13: 171-183.
- (1986). On the relationship between two fatigue-life models. IEEE transactions on reliability 35: 167-169.
- Díaz García, J.A., Leiva Sánchez, V. (2002). A new family of life distributions based on Birnbaum-Saunders distribution. Reporte Técnico CIMAT 03.09.2002,I-02-17(PE), Centro de Investigaciones Matemáticas, Guanajuato, México. (<http://www.cimat.mx/biblioteca/RepTec>).
- Engelhardt, M., Bain, L.J. and Wright, F.T. (1981). Inferences on the parameters of the Birnbaum-Saunders fatigue life distribution based on maximum likelihood estimation. Technometrics 23: 251-256.

- 9 McCarter, K.S. (1999). Estimation and prediction for the Birnbaum-Saunders distribution using Type-II censored samples, with a comparison to the Inverse Gaussian Distribution. Unpublished Ph. D. Thesis. Kansas State University, Department of Statistics, College of Arts and Sciences.
- 10 Owen, W.J., Padgett, W.J. (1999). Accelerated test models for system strength based on Birnbaum-Saunders distribution. Kluwer academic publishers. Lifetime data analysis 5: 133-147. Boston.
- 11 Rieck, J.R. (1995). Parameter estimation for the Birnbaum-Saunders distribution based on symmetrically censored samples. Commun. Statist. Theory Meth. 24 (7): 1721-1736.
- 12 Rieck, J.R., Nedelman J. (1991). A log-linear model for the Birnbaum-Saunders distribution. Technometrics 33: 51-60,



## Area de Fisiotecnia Vegetal

Tels. (844) 411-02-96 y 411-02-97 · [fborrego@uaaan.mx](mailto:fborrego@uaaan.mx)

El Area de Fisiotecnia Vegetal del Departamento de Fitomejoramiento de la UAAAN cuenta con Laboratorio e Invernadero en el Campus Buenavista, y lotes en campos experimentales y con agricultores cooperantes, en donde se realiza investigación tecnológica para generar fitotipos (plantas modelo) de variedades mejoradas con mayor eficiencia metabólica, y con mayor cantidad y calidad de rendimiento en cultivos con adaptación a factores adversos, para agricultura intensiva y sustentable en zonas áridas y semiáridas.

Actualmente se realizan proyectos en los cultivos de nopal, tomate, melón y papa, se brinda apoyo en estudios específicos de tolerancia a sequía en trigo y maíz, y se realizan análisis de crecimiento en cilantro y calabacita, así como estudios de estomas y clorofila en pino, eficiencia fotosintética en chile y vid, e incremento en calidad alimenticia en nopal para forraje. Se ha trabajado en convenios con CONACYT, FONCAN, La Forestal F.C.L. y Nevada Chemicals, para proyectos específicos con candelilla, nopal, orégano y en eficiencia fisiológica de ácidos húmicos en melón.

Se cuenta con instrumental científico avanzado y de alta precisión, como fotosintetómetro portátil (LI-6200), transpirómetro portátil (LI-1600), medidor portátil de Índice de Area Foliar (LAI-2000), sensores de radiación, temperatura y humedad, medidor electrónico de área foliar (LI-3000), cámaras bioclimáticas, estufas de secado, etc. Se realizan determinaciones de clorofila, ácido ascórbico, estomas, vitaminas, etc.

Asimismo se imparte docencia a Licenciatura, Maestría y Doctorado, así como Talleres y Cursos de Educación Continua a nivel nacional e internacional sobre aprovechamiento integral de nopal, a través de la FAO.

### Personal

**Coordinador del Area:** Dr. Fernando Borrego Escalante

**Jefa de Laboratorio:** Dra. Ma. Margarita Murillo Soto

**Docentes:** M.C. Ma. Cristina Vega Sánchez, M.C. Arnoldo Oyervides García, Ing. José Luis Guerrero Ortiz





## Centro de Investigación en Reproducción Caprina

Periférico y Carretera a Santa Fe, A.P. 940 Torreón, Coahuila, México

Tel: (871) 733-1270, 733-1090 . Fax: (871) 733-1210

Responsable: Dr. José Alberto Delgadillo Sánchez . [delgadil@ul.uaaan.mx](mailto:delgadil@ul.uaaan.mx)

### Quienes somos

El Centro de Investigación en Reproducción Caprina (CIRCA) es un espacio académico que, transfiere la tecnología que genera y ofrece asistencia técnica integral a los caprinocultores desde 1992, y en el que interactúan docentes, alumnos y productores. El Centro brinda el ambiente ideal para que los interesados -escolarizados o no- aprendan, en la práctica, técnicas y manejo reproductivos de los caprinos.

### Objetivo general

Realizar investigación científica y tecnológica sobre el comportamiento reproductivo de los caprinos de la Comarca Lagunera (México) en particular, y el de las zonas subtropicales en general, que permita la vinculación de la UAAAN con el sector productivo, e interactuar con instituciones nacionales e internacionales relacionadas con la biología de la reproducción animal, así como apoyar la formación de alumnos de licenciatura y postgrado.

### Objetivos específicos

- Determinar las características reproductivas de los caprinos de la Comarca Lagunera y su interacción con el medio ambiente.
- Implementar técnicas de control reproductivo acordes con los sistemas de explotación.

En la actualidad el CIRCA colabora en aspectos académicos, científicos y tecnológicos con las siguientes instituciones:

- Estación de Fisiología de la Reproducción y del Comportamiento. INRA de Nouzilly, Francia.
- Centro Nacional de Neurobiología. UNAM, Juriquilla, Querétaro.
- Centro de Investigación en Reproducción Animal. CINVESTAV-U.A.T.
- caprinocultores de la Comarca Lagunera (México).

Debido a la relevancia de su quehacer y de los resultados obtenidos hasta la fecha, el CIRCA recibe apoyos económicos de fuentes de financiamiento externas a la Universidad, tales como:

- Consejo Nacional de Ciencia y Tecnología.
- Fundación Internacional para la Ciencia, Estocolmo, Suecia.
- Sistema Regional de Investigación Alfonso Reyes. ANUIES-SEP-CONACyTIECOS.

### Reconocimientos

- Fundación Internacional para la Ciencia, 2001 (Premio por investigación)
- Consejo Estatal de Ciencia y Tecnología de Coahuila, 2002 (Reconocimiento por investigación).

### Investigaciones en proceso

- Determinación de los mecanismos fotoperiódicos responsables del ciclo anual de reproducción de los caprinos de la Comarca Lagunera.
- Determinación de las interacciones entre la tasa de ovulación, la mortalidad embrionaria, los abortos y la fertilidad de las cabras y los sistemas de explotación
- Determinación del efecto del sistema de explotación sobre el ciclo anual de reproducción de los machos **cabríos**, así como la respuesta de éstos a los tratamientos fotoperiódicos para inducir la actividad sexual.
- Identificación de los mecanismos involucrados en la respuesta de las hembras al efecto macho (comportamiento sexual, olor, vocalizaciones..).
- Manipulación de la relación madre-cría, para incrementar la producción de leche y disminuir el anestro postparto.
- Determinación de las causas de la mortalidad perinatal de los cabritos explotados extensivamente.



# Indice de Funcionamiento de Normas de Paro en Control Estadístico de Calidad

Juan Manuel Saucedo Esquivel

Departamento de Estadística y Cálculo, Universidad Autónoma Agraria Antonio Narro, Buenavista, Saltillo Coah., México.

---

**Abstract.** *Performance Index of Stopping Rules in Statistical Control of Quality.* This work deals with the idea of stopping rules, an instrument used like a means of securing the quality of end products in a manufacturing process. Its main purpose is to formulate a general stopping rule, and to obtain an expression for its index of operation. This objective is reached by applying some ideas of dynamic programming that will allow an approach to the problem from an algebraic perspective.

Key Words: Control Charts, Dynamic Programming, Conditional Expectation, Linear Equations.

**Resumen.** Este trabajo trata sobre la idea de norma de paro, instrumento que se utiliza como medio de aseguramiento de la calidad de los productos finales en un proceso de fabricación. El propósito central es formular una regla general de paro y obtener una expresión para su índice de funcionamiento, objetivo que se alcanza aplicando ideas de programación dinámica que permiten abordar el problema desde una perspectiva algebraica.

**Palabras Clave:** Gráficos de control, Regla de paro, Programación dinámica Esperanza condicional, Ecuaciones lineales.

---

## Introducción

Este trabajo trata sobre un problema fundamental que se presenta en cualquier proceso productivo, esto es, asegurar la calidad de los artículos finales. Aunque la noción en calidad es una idea compleja, este concepto puede definirse, desde una perspectiva operacional, como la aptitud para el uso que los productos tendrán, de tal forma que un propósito central en cualquier proceso de fabricación es lograr que las características esenciales de los objetos producidos, las cuales determinan la adecuación para el empleo que se les dará, permanezcan constantes, asegurando que la calidad de los productos finales que llegan al mercado sea similar. Cuando los rasgos esenciales de los artículos fabricados permanecen inalterados conforme se desarrolla la producción, se dice que el proceso productivo se encuentra bajo control estadístico, mientras que si uno o más de los rasgos relevantes se altera, el proceso se encuentra fuera de control. En términos generales, la estrategia que se sigue para detectar cambios no deseados en los rasgos de interés puede describirse como sigue ( Godfrey y Mundel (1984)): A intervalos regulares (i) se toman muestras de artículos finales, (ii) se determina la característica de interés en cada uno de ellos, y (iii) los datos obtenidos se utilizan de

cierta forma para detectar cambios no deseados en los rasgos relevantes; en este último punto la información recabada en las inspecciones muestrales sucesivas se traslada al área de toma de decisiones. Cuando se juzga que los datos acumulados indican que el proceso se encuentra fuera de control, el proceso productivo se revisa, buscando las causas que han propiciado la alteración de los rasgos esenciales; debido a que dicha revisión del proceso de fabricación puede implicar que se detenga la producción, en donde la regla que lleva a declarar al proceso como fuera de control se conoce como regla o norma de paro (Montgomery (1985), Duncan (1989) y John (1990)). Desde luego, los datos obtenidos en las inspecciones sucesivas pueden emplearse de muy diversas formas para declarar que el proceso se encuentra fuera de control, esto es, existen múltiples normas de paro que pueden utilizarse para detectar cambios en los rasgos relevantes de los artículos finales.

El objetivo de este trabajo es estudiar una norma general de paro y determinar su índice de funcionamiento. Dicho indicador proporciona una idea, a priori, acerca del número de productos finales que se fabricarán antes de que una alteración en la calidad de los mismos sea detectada aplicando una regla de paro. El resultado

principal que se obtiene en esta nota está contenido en el Teorema 4.1, el cual establece una fórmula para el número esperado de inspecciones que se realizarán antes de detectar que el proceso se encuentre fuera de control, proporcionando una visión unificante de diversos resultados establecidos en la literatura (John (1990)). El enfoque que se utiliza depende de argumentos que involucran esperanzas condicionales. Estas ideas han sido exitosamente aplicadas en problemas de programación dinámica, los cuales se refieren a procesos secuenciales de toma de decisiones; Vea, por ejemplo, Sennott (1995, 1998), Puterman (1994), Cavazos-Cadena (1991, 1996).

El presente trabajo ha sido organizado de la siguiente manera: En la Sección 2 se describe el instrumento básico en el área de control de calidad, a saber, la idea de gráfico de control, cuya implementación permite organizar de manera visual la información que se recaba en las inspecciones muestrales que periódicamente se realizan en la línea de producción; de manera paralela, se introduce la noción de norma de paro, a través de la cual la información que se acumula en el gráfico se traduce a decisiones concretas de revisar el proceso productivo, o de continuarlo normalmente. En la Sección 3 se define el índice de funcionamiento de una norma de paro, el cual se construye tomando en cuenta el número de inspecciones que transcurrirán infructuosamente antes de detectar alteraciones del valor promedio del rasgo relevante de los artículos finales. A continuación, la Sección 4 contiene la formulación de la norma de paro general que es el objeto de estudio de este trabajo, estableciendo el resultado principal en el Teorema 4.1, el cual se demuestra en la Sección 6 después de presentar los resultados técnicos preliminares en la Sección 5. Finalmente, la presentación culmina en la Sección 6 con conclusiones y comentarios breves antes de la literatura citada.

### Gráficos de Control y Normas de Paro

Un gráfico de control es un medio para organizar y representar visualmente la información obtenida al inspeccionar periódicamente muestras de productos finales. Este instrumento fue introducido por Shewart (1931), y desde sus orígenes se ha mantenido como la herramienta esencial para asegurar el control estadístico de un proceso productivo. Como punto de partida, sea  $q$  el valor de un rasgo de interés determinado y note que el objetivo básico es detectar cualquier cambio en  $q$  conforme el proceso productivo se desarrolla, esto es, se desea mantener  $q$  igual a su valor original  $\bar{q}$ , este último puede pensarse como el valor de diseño del rasgo relevante, el cual asegura que los artículos fabricados

satisfagan, en una gran proporción, los requerimientos impuestos por el consumidor.

Para asegurar que el valor actual de  $q$  coincida con su nivel original  $\bar{q}$ , periódicamente se analizan muestras de productos y se determina el valor del rasgo relevante en cada uno de los artículos en la muestra. En el desarrollo de este trabajo, se supondrá que cada una de las muestras extraídas es de tamaño  $n$  y las observaciones obtenidas se representaran mediante  $X_1, X_2, \dots, X_n$ . A partir de estos números se construye el estadístico

$$Y = \frac{X_1 + X_2 + \dots + X_n}{n} \tag{2.1}$$

la media muestral de  $X_1, X_2, \dots, X_n$ .

**Construcción de un Gráfico de Control.** Con la notación dada en (2.1), los pasos para construir un gráfico de control asociado al nivel  $\bar{q}$ , del rasgo relevante, pueden describirse como sigue (Duncan (1989), John (1990)):

**Etap 1.** Determinar o postular la distribución de  $Y$  cuando el parámetro  $q$  asume su valor original  $\bar{q}$ .

**Etap 2.** Determinar una región  $\bar{A}$  tal que

$$P\{Y \in \bar{A}\} = 1 - \alpha \tag{P.2}$$

Esta igualdad significa que, bajo la condición de que  $q$  asume su valor original  $\bar{q}$ , la probabilidad de que  $Y$  tome un valor en la zona  $\bar{A}$  es  $1 - \alpha$ : este último número es el nivel de inclusión de la región  $\bar{A}$ , la cual se conoce como zona de predicción para  $Y$  con nivel  $1 - \alpha$  (Graybill (1986), Capítulo 5). Al seleccionar un  $1 - \alpha$  cercano a uno, entonces es 'prácticamente seguro' que la inclusión  $Y \in \bar{A}$  se satisfaga. Usualmente, se selecciona  $\bar{A}$  como un intervalo  $[L, U]$ , donde

$$P\{L \leq Y \leq U\} = 1 - \alpha \tag{2.3}$$

y en este caso  $L$  y  $U$  se conocen los límites de control inferior y superior, respectivamente.

**Etap 3.** Marcar la zona  $\bar{A}$  sobre el eje vertical del plano coordenado y delimitar la región

$$A = \{(x, y) | y \in \bar{A}\}$$

la cual se denomina como la zona de aceptación.

**Etap 4.** De los productos finales seleccione, a intervalos posiblemente regulares, bloques disjuntos de  $n$  artículos, calcule el valor  $Y_i$  de la media de la característica de interés de los  $n$  productos finales seleccionados en el  $i$ -ésimo grupo y marque el punto correspondiente  $(i, Y_i)$  en el plano coordenado; usualmente el tamaño  $n$  de los

bloques inspeccionados es un número pequeño, como por ejemplo cuatro o cinco. El gráfico de control es la zona de aceptación  $A$ , junto con las marcas que sucesivamente se van incorporando al plano coordenado al realizar las inspecciones.

**El Empleo Básico de un Gráfico de Control.** Para entender la aplicación de un gráfico de control, observe primero que para cada dato  $Y_i$  se tiene la siguiente equivalencia:

$$Y_i \in \bar{A} \Leftrightarrow (i, Y_i) \in A$$

Por otro lado, debido a que el nivel de inclusión  $1 - \alpha$  se selecciona 'cerca de 1', el observador percibe como 'prácticamente seguro' que cuando el valor actual de  $\bar{Y}$  es  $\bar{y}$ , cada dato  $Y_i$  pertenezca a  $\bar{A}$ , lo cual equivale a que los puntos  $(i, Y_i)$  que se marcan en el plano coordenado pertenezcan a la zona de aceptación  $\bar{A}$ . En otras palabras, el observador considerará 'dentro de lo normal' el hecho de que los puntos que sucesivamente se señalan en el plano coordenado se ubiquen dentro de  $A$  y pensará que el valor de  $\bar{Y}$  se mantiene en el nivel original  $\bar{y}$ . Sin embargo, cuando una de esas marcas se localice fuera de la región  $A$ , habrá observado un evento que, bajo la condición de que  $\bar{Y} = \bar{y}$ , tiene sólo una probabilidad  $\alpha$  de ocurrir, percibiendo este hecho como una señal de que el valor de  $\bar{Y}$  se ha modificado, alejándose del valor  $\bar{y}$ , que tenía originalmente, de manera que el valor medio de la característica de interés se ha alterado. En estas circunstancias, se emprenderán acciones que busquen detectar la causa del cambio en el valor de  $\bar{Y}$  y se tomarán las medidas correctivas necesarias. La norma de actuación que se ha descrito es un ejemplo de regla de paro, pues la revisión del proceso de producción puede, en general, implicar un paro parcial o total de las actividades. Desde luego, ésta no es la única regla de paro que puede derivarse de un gráfico de control y, para describir otras reglas de uso común, es conveniente introducir una componente auxiliar de un gráfico.

**La Franja de Advertencia.** Además de la zona de aceptación de  $A$  descrita anteriormente, en un gráfico también se incluye, generalmente, una zona denominada de advertencia, la cual se denota mediante  $W$ , y que está formada por puntos en el plano que son cercanos a la frontera de la región de aceptación; la forma específica de  $W$  depende de la aplicación que se tenga en mente; esta idea se ilustrará por medio de un ejemplo más adelante. En este punto es conveniente mencionar que cuando una marca obtenida de una media muestral se ubica en  $w$ , el observador no declara de inmediato que el proceso se encuentra fuera de control, sino que se emite

una señal de alerta que será tomada en cuenta, conjuntamente con las futuras observaciones, para establecer la necesidad de revisar el proceso productivo. **Norma de Paro.** Formalmente, una regla o norma de paro se obtiene especificando uno o más patrones de las marcas trazadas sucesivamente en el gráfico de control, de manera que cuando uno de dichos patrones es observado, se declara que el valor actual del parámetro de interés no coincide con su nivel original y se emite la orden de revisar el proceso de fabricación. Esta es una definición bastante general y es claro que existen varias de normas de paro posibles.

Los conceptos anteriores se ilustran en el siguiente ejemplo.

**Ejemplo 2.1.** El gráfico de control más común en las aplicaciones se obtiene estableciendo las siguientes expresiones para los límites de control  $L$  y  $U$  en (2.3):

$$L = \bar{y} - 3\tau_0, \quad U = \bar{y} + 3\tau_0 \quad (2.4)$$

Especificación que se conoce como la regla de las tres desviaciones estándar; en estas igualdades,  $\tau_0^2$  es la varianza poblacional que se obtiene a partir de los datos obtenidos en la  $i$ -ésima inspección. Note que en este caso, la región  $\bar{A}$  en la cual el observador considera como prácticamente seguro que las medias muestrales se ubiquen, es  $A = [i, Y_i] \in [i, \bar{y} - 3\tau_0, \bar{y} + 3\tau_0]$  y que el correspondiente nivel de inclusión esta dado por

$$1 - \alpha = P\left\{ \bar{Y} \in [i, \bar{y} - 3\tau_0, \bar{y} + 3\tau_0] \right\};$$

bajo el supuesto de que  $Y_i \sim N(\theta_0, \tau_0^2)$ , se tiene que

$$1 - \alpha = P_{\theta_0} [Y \in A] = 0.99730.$$

Por otro lado, note que la zona de aceptación  $A$  en el plano coordenado es

$$A = \{ (x, y) \mid (x, y) \in \theta_0 - 3\tau_0, \bar{y} \leq y \leq \bar{y} + 3\tau_0 \}$$

En este momento se está en condiciones de enunciar la regla de paro más simple:

**RP.** Declarar el proceso como fuera de control cuando un punto  $(i, Y_i)$  se ubique fuera de  $A$ .

De acuerdo a esta norma de paro, se declara que el parámetro de interés ha alterado su valor original cuando una media observada se desvíe de  $\bar{y}$ , por más de tres desviaciones estándar. Por supuesto, esta norma no es la única que es posible enunciar. Considere, por ejemplo, la siguiente:

**RP.** Declarar el proceso como fuera de control cuando se acumulen dos marcas, no necesariamente consecutivas, fuera de  $A$ .

Otras reglas comúnmente utilizadas involucran a la región de advertencia. Como ya se ha mencionado, dicha zona esta formada por puntos 'cercanos' a la frontera de

la zona de aceptación. En el caso que nos ocupa, es costumbre definir

$$W = \{(x, y) | \theta_0 - 3\tau_0 \leq y \leq \theta_0 - 2\tau_0\} \cup \{(x, y) | \theta_0 + 3\tau_0 \leq y \leq \theta_0 + 2\tau_0\},$$

y la siguiente regla utiliza esta zona para decidir cuándo debe revisarse el proceso productivo.

**RP<sub>11</sub>**: Declarar el proceso como fuera de control cuando se acumulen dos marcas consecutivas en la zona  $W \cup A^c$ , la unión de la región de advertencia W y el complemento de la zona de aceptación A.

### El Índice de Funcionamiento de una Norma de Paro

En general, el propósito de una norma de paro es evitar que se produzcan artículos cuyo rasgo relevante se aleje del valor original  $\theta_0$ . Debido a que los artículos finales son analizados solamente durante las inspecciones realizadas, una regla de paro será mejor en la medida en que detecta un cambio en la característica de interés al tomar un número menor de inspecciones. Para definir el índice de funcionamiento de manera formal, es conveniente introducir primero la siguiente noción auxiliar.

**Definición 3.1** Considere una regla de paro determinada y suponga que al seleccionar la k-ésima muestra el valor esperado del rasgo de interés es  $\theta_0$ , mientras que al analizar la muestra k + 1, dicho valor es  $\theta^*$ , el cual ya no sufre alteración posterior. Sea k + N el instante de inspección en que, de acuerdo a la regla de paro estipulada, el observador declara que es necesario revisar el proceso productivo.

En este caso, la sucesión de medias

$$Y_{k+1}, Y_{k+2}, \dots, Y_{k+N},$$

las cuales se observan desde que  $\theta$  altera su valor inicial hasta que se declara la necesidad de inspeccionar el proceso, se llama una corrida de datos, mientras que N es la longitud de la corrida.

Es importante notar que N, la longitud de la corrida en la definición anterior, es el número de inspecciones que se realizan en la línea de producción antes de que el proceso se revise. Cuando las muestras se toman a intervalos regulares y la tasa de producción es constante, N es proporcional al número de artículos producidos desde que el valor del rasgo de interés se alteró su valor, y el fabricante está naturalmente interesado en que N sea lo más pequeño posible. De esta forma, ante la posibilidad de elegir entre dos reglas de paro, se preferirá aquella que produzca menores longitudes de corridas, esto es, los valores de N pueden utilizarse para medir el desempeño de una norma de actuación. Sin embargo, N es una variable aleatoria y como índice de funcionamiento de una regla de paro se utiliza el valor esperado de N.

**Definición 3.2.** En el contexto de la Definición 3.1, la longitud promedio de la corrida se define mediante

$$LPC = LPC(\theta^*) = E_c [N], \tag{3.1}$$

y  $LPC(\theta^*)$  es el índice de funcionamiento de la norma de paro considerada.

Antes de ejemplificar la determinación del índice de desempeño de una norma de paro, es oportuno describir el procedimiento general: En el momento en que se extrae la k-ésima muestra de n productos, el parámetro en cuestión conserva su valor original  $\theta_0$ , pero al obtener la muestra k+1 el parámetro asume ya otro valor  $\theta^*$ . Este cambio no es percibido de forma inmediata y el proceso productivo continúa hasta que, al seleccionar una muestra en la ocasión k+N, el inspector emite la señal de alerta declarando que el valor de  $\theta$  ya no es el original. El valor de N es aleatorio y su distribución depende de la regla de paro que se utilice. De acuerdo a la Definición 3.2, la longitud promedio de la corrida es

$$LPC(\theta^*) = E_c [N] = \sum_{n=1}^m n P_{\theta^*} [N = n]. \tag{3.2}$$

así que un método para encontrar  $LPC(\theta^*)$  consiste en determinar la distribución de N cuando el parámetro asume el valor  $\theta^*$  y utilizar la expresión (3.2) para calcular la esperanza de N. Este enfoque se ilustra a continuación.

**Ejemplo 3.1.** (i) Considere la regla de paro  $RP_1$  introducida en el Ejemplo 2.1. De acuerdo a esta norma el proceso se declara fuera de control tan pronto como una marca se ubique fuera de la zona  $A$  dada en (2.5). Para determinar el índice de funcionamiento de  $RP_1$ , suponga que en la  $k$ -ésima inspección el valor esperado del rasgo relevante es su valor original  $\theta^*$ , pero que al momento de extraer la muestra  $k+1$  dicho valor se ha alterado, asumiendo el valor  $\theta^*$ , el cual se mantiene fijo a partir de ese instante. Suponiendo que las medias  $Y_i$  obtenidas de las inspecciones sucesivas son independientes con distribución común  $N(\theta^*, \tau)$ , se tiene que

$$\begin{aligned}
 P_{\theta^*} [N = n] &= P_{\theta^*} [Y_{k+1} \in A, Y_{k+2} \in A, \dots, Y_{k+n-1} \in A, Y_{k+n} \notin A] \\
 &= P_{\theta^*} [Y_{k+1} \in A] \times P_{\theta^*} [Y_{k+2} \in A] \times \dots \times P_{\theta^*} [Y_{k+n-1} \in A] \times P_{\theta^*} [Y_{k+n} \notin A] \\
 &= P_{\theta^*} [Y_{k+1} \in A]^{n-1} \times (1 - P_{\theta^*} [Y_{k+1} \in A]) \\
 &= (1 - p(\theta^*))^{n-1} p(\theta^*)
 \end{aligned} \tag{3.3}$$

donde  $p(\theta^*)$  está dado por

$$\begin{aligned}
 p(\theta^*) &= 1 - P_{\theta^*} [Y_{k+1} \in A] \\
 &= 1 - P_{\theta^*} [e - 3\tau_0 \leq Y_{k+1} \leq e + 3\tau_0] \\
 &= 1 - P_{\theta^*} \left[ \frac{\theta_0 - \theta^*}{\tau_0} - 3 \leq \frac{Y_{k+1} - \theta^*}{\tau_0} \leq \frac{\theta_0 - \theta^*}{\tau_0} + 3 \right] \\
 &= 1 - \Phi \left( \frac{\theta_0 - \theta^*}{\tau_0} + 3 \right) + \Phi \left( \frac{\theta_0 - \theta^*}{\tau_0} - 3 \right) \\
 &= \Phi \left( -\frac{\theta_0 - \theta^*}{\tau_0} - 3 \right) + \Phi \left( \frac{\theta_0 - \theta^*}{\tau_0} - 3 \right)
 \end{aligned} \tag{3.4}$$

Donde  $\Phi(\cdot)$  es la función de distribución Normal estándar. De acuerdo a (3.3), la longitud de la corrida  $N$  que tiene una distribución Geométrica con parámetro  $p(\theta^*)$ , de manera que su valor esperado es

$$E_{\theta^*} [N] = \sum_{n=1}^{\infty} n P_{\theta^*} [N = n] = \frac{1}{p(\theta^*)}$$

(ii) Para la norma  $RP_2$ , descrita en el Ejemplo 2.1, no es difícil ver que  $N$  tiene distribución Binomial negativa con parámetros  $2$  y  $p(\theta^*)$ , de manera que

$$E_{\theta^*} [N] = \sum_{n=1}^{\infty} n P_{\theta^*} [N = n] = \frac{1}{p(\theta^*)}$$

En el ejemplo anterior, el índice de funcionamiento de las normas consideradas se calculó determinando la distribución exacta de la longitud  $N$  de una corrida. Sin embargo, para reglas de paro más complejas, este enfoque no es aplicable, pues dado el enorme número de reglas de paro potenciales, no es de esperarse que calcular la distribución exacta de  $N$  para una norma arbitraria sea un problema susceptible de tener una solución simple. Por esta razón, es interesante explorar la posibilidad de obtener el indicador de funcionamiento de una norma a través de métodos indirectos, esto es, que no involucren la distribución exacta de  $N$ . Las ideas que se emplearan en esta dirección se ilustran en el siguiente teorema y han sido adaptadas de argumentos utilizados en problemas de programación dinámica,

en donde la propiedad de Markov permite emplear esperanzas condicionales de manera simple y fructífera (Sennot (19%), Puterman (1994), Cavazos-Cadena (1991, 1996)). A continuación se determina el índice de funcionamiento de la regla  $RP_{wA}^c$  introducida en el Ejemplo 2.1. Para establecer el resultado en forma precisa, defina

$$A_2 = \{(x, y) | \theta_0 - 2\tau_0 \leq y \leq \theta_0 + 2\tau_0\}$$

$$p_{A_2} = p_{A_2}(\theta^*) = P_{\theta^*} \left[ (i, Y_i) \in A_2 \right] \tag{3.5}$$

y observe que

$$W \cup A^c = A_2^c$$

de tal forma que de acuerdo a  $RP_{wA}^c$ , se emite la señal de revisión del proceso productivo cuando dos marcas consecutivas se ubican fuera de  $A_2$ .

**Teorema 3.1** Para la regla  $RP_{wA}^c$ , la longitud promedio de una corrida es

$$LPC(\theta^*) = \frac{1}{1 - p_{A_2}(\theta^*)} + \frac{1}{(1 - p_{A_2}(\theta^*))^2}$$

donde  $p_{A_2}(\theta^*)$  está dada por (3.5).

**Demostración.** Como punto de partida, es conveniente introducir la siguiente notación:  $C_k$  denota 'el estado' en el que se tienen k puntos consecutivos fuera de  $A_2$ , desde el último momento en que se observó un punto dentro de  $A_2$ , y  $E_{\theta^*} [N|C_k]$  es la esperanza condicional de N dado que se tienen acumuladas k observaciones fuera de  $A_2$ . Observe que

$$LPC(\theta^*) = E_{\theta^*} [N|C_0]$$

pues una corrida se cuenta a partir de un momento 'normal' en el que el parámetro  $\theta$  conserva su valor original  $\theta^*$ . Además,

$$E_{\theta^*} [N|C_1] = 0,$$

pues al acumular dos datos consecutivos fuera de  $A_2$ , se emite la señal de revisar el proceso. Considere ahora los siguientes casos que pueden presentarse en el momento de seleccionar las muestras conforme se desarrolla el proceso:

**Caso 1.** Se tienen acumuladas cero observaciones fuera de la franja  $A_2$ , esto es, se está observando el estado  $C_0$ . En esta situación, se extrae la muestra y se calcula su media. Esta última puede caer dentro de  $A_2$ , con probabilidad  $p_{A_2}(\theta^*)$ , aumentando en 1 la longitud de la corrida y dejando en cero el número de observaciones acumuladas fuera de  $A_2$ , de manera que al realizar la siguiente inspección se enfrentara la misma situación. Por otro lado, la media observada puede caer fuera de  $A_2$ , con probabilidad  $1 - p_{A_2}(\theta^*)$ , incrementando en uno la longitud de la corrida y provocando que la siguiente inspección se realice teniendo una observación acumulada fuera de  $A_2$ , esto es, ante el estado  $C_1$ . Luego,

$$\begin{aligned} E_{\theta^*} [N|C_0] &= 1 + E_{\theta^*} [N|C_0] p_{A_2}(\theta^*) + (1 + E_{\theta^*} [N|C_1]) (1 - p_{A_2}(\theta^*)) \\ &= 1 + E_{\theta^*} [N|C_0] p_{A_2}(\theta^*) + E_{\theta^*} [N|C_1] (1 - p_{A_2}(\theta^*)) \end{aligned} \tag{3.6}$$

**Caso 2** Ahora suponga que se inicia la inspección teniendo una observación acumulada fuera de la franja  $A_2$ , esto es, ante el estado  $C_1$ . En este caso, se extrae la muestra y se calcula su media; si ésta se ubica dentro de  $A_2$ , entonces se incrementa en uno la longitud de la corrida y en la siguiente extracción se estará ante el estado  $C_0$ . Por otro lado, si la

media observada cae fuera de  $A$ , entonces se aumenta en uno la longitud de la corrida, pero se declara la señal de revisión, pues ya se tendrán acumulados dos puntos fuera de la franja de advertencia  $A$ . Por lo tanto

$$\begin{aligned} E_{\theta^*}[N|C_1] &= (1 + E_{\theta^*}[N|C_0])p_{A_2}(\theta^*) + 1 \times (1 - p_{A_2}(\theta^*)) \\ &= 1 + E_{\theta^*}[N|C_0]p_{A_2}(\theta^*) \end{aligned} \tag{3.7}$$

A partir de estas igualdades puede despejarse  $E[N|C_0]$ . Con este propósito, observe que (3.6) puede escribirse equivalentemente como

$$E_{\theta^*}[N|C_0](1 - p_{A_2}(\theta^*)) = 1 + E_{\theta^*}[N|C_1](1 - p_{A_2}(\theta^*))$$

y reemplazando la expresión para  $E[N|C_1]$  obtenida en (3.7) se obtiene que

$$\begin{aligned} E_{\theta^*}[N|C_0](1 - p_{A_2}(\theta^*)) &= 1 + (1 + E_{\theta^*}[N|C_0]p_{A_2}(\theta^*))(1 - p_{A_2}(\theta^*)) \\ &= 1 + (1 - p_{A_2}(\theta^*)) + E_{\theta^*}[N|C_0]p_{A_2}(\theta^*)(1 - p_{A_2}(\theta^*)) \end{aligned}$$

y entonces, finalmente

$$E_{\theta^*}[N|C_0] = \frac{2 - p_{A_2}(\theta^*)}{(1 - p_{A_2}(\theta^*))^2} = \frac{1}{1 - p_{A_2}(\theta^*)} + \frac{1}{(1 - p_{A_2}(\theta^*))^2}, \tag{3.8}$$

estableciendo la conclusión deseada.

**Ejemplo 3.2.** Suponga que las medias  $\theta_0, \dots, \theta_{k+1}$  son independientes con distribución común  $N(\theta^*, \tau^2)$ . A partir de la definición de  $p_{A_2}(\theta^*)$ , esta cantidad puede expresarse en términos de la distribución Normal estándar:

$$\begin{aligned} p_{A_2}(\theta^*) &= 1 - P_{\theta^*}[Y_i \in [e_0 - 2\tau, e_0 + 2\tau]] \\ &= 1 - P_{\theta^*}\left[-2 + \frac{\theta_0 - \theta^*}{\tau_0} < \frac{Y_i - \theta^*}{\tau_0} < 2 + \frac{\theta_0 - \theta^*}{\tau_0}\right] \\ &= 1 - \left[ \Phi\left(2 + \frac{\theta_0 - \theta^*}{\tau_0}\right) - \Phi\left(-2 + \frac{\theta_0 - \theta^*}{\tau_0}\right) \right] \end{aligned}$$

Luego, si  $\theta^* = \theta_0 + c\tau_0$ , entonces

$$\begin{aligned} p_{A_2}(\theta^*) &= 1 - \left[ \Phi\left(2 + \frac{\theta_0 - \theta^*}{\tau_0}\right) - \Phi\left(-2 + \frac{\theta_0 - \theta^*}{\tau_0}\right) \right] \\ &= 1 - [\Phi(2 - c) - \Phi(-2 - c)] \\ &= [1 - \Phi(2 - c)] + \Phi(-2 - c) \\ &= \Phi(c - 2) + \Phi(-2 - c). \end{aligned}$$

Por ejemplo, si  $\theta^* = \theta_0 + c\tau_0$ , entonces  $p_{A_2}(\theta^*) = \Phi(-1) + \Phi(-3) = 0.15866 + 0.00270 = 0.16136$ . Luego,

$$LPC(\theta_0 + \tau_0) = \frac{1}{1 - .16136} + \frac{1}{(1 - .16136)^2} = 2.614241$$

Similarmenle, si  $\theta^* = \theta_0 + 2\tau_0$ , entonces  $P_{1,m}(\theta^*) = \Phi(0) + \Phi(4) \approx 0.50000$ , de manera que

$$LPC(\theta_0 + 2\tau_0) \approx \frac{1}{1 - 0.50000} + \frac{1}{(1 - 0.50000)^2} = 6.$$

### La Norma $RP_{1,m}$ y el Resultado Principal

En esta sección se determina la longitud promedio de una corrida para una regla de paro generalizada. Para describir esta regla de toma de decisiones, sea  $A$  una zona en el plano coordinado en la cual se considera 'normal' que se ubiquen las marcas sucesivas que se señalan en la Sección 2. El complemento de  $A$  se denota mediante  $R^*$  y se refiere como la zona de rechazo, es decir,

$$R^* = A^c$$

Por otro lado, suponga que se ha delimitado una región de advertencia  $W^*$  de tal manera que

$$W^* \subset A,$$

y defina la zona de aceptación absoluta mediante

$$A^* = A \setminus W^* = \{(x, y) | (x, y) \in A \text{ and } (x, y) \notin W^*\}$$

La norma de paro  $RP_{1,m}$  se describe a continuación:

**$RP_{1,m}$** : El proceso se declara fuera de control estadístico si se encuentra un punto en la zona de rechazo  $R^*$  o cuando se tienen  $m$  puntos consecutivos en la región de advertencia  $W^*$ .

Como en la sección precedente, sea  $\zeta_k$  el estado en el cual hay  $k$  puntos acumulados consecutivamente en la zona de advertencia y observe que mediante un argumento de esperanza condicional similar al utilizado en la demostración del Teorema 3.1, se obtiene que para  $k = 1, 2, \dots, m - 1$ ,

$$E_{\theta_0}[N|C_k] = 1 \times P_{R^*}(\theta) + (1 + E_{\theta_0}[N|C_0])P_{A^*}(\theta) + (1 + E_{\theta_0}[N|C_{k+1}])P_{W^*}(\theta) \tag{4.1}$$

Note que en este caso

$$E_{\theta_0}[N|C_m] = 0,$$

pues la regla emite la señal de revisión cuando se tienen  $m$  puntos consecutivos en la zona de advertencia. Observe, además, que

$$P_{A^*}(\theta) + P_{W^*}(\theta) + P_{R^*}(\theta) = 1. \tag{4.2}$$

El siguiente teorema, en el cual se determina  $E[\zeta_k]$  para todo  $k = 1, 2, \dots, m - 1$ , es el principal resultado de este trabajo.

**Teorema 4.1.** Bajo la regla  $RP_{1,m}$  las siguientes conclusiones (i) y (ii) son válidas.

(i) La longitud promedio de una corrida está dada por

$$LPC(\theta) = E_{\theta}[N|C_0] = \frac{1 - p_w(\theta)^m}{1 - p_s(\theta) - p_w(\theta) + p_s(\theta)p_w(\theta)^m} \tag{4.3}$$

Más aún,

(ii) Para  $k = 1, 2, \dots, m - 1$

$$E_{\theta}[N|C_k] = \frac{1 - p_w(\theta)^{m-k}}{1 - p_s(\theta) - p_w(\theta) + p_s(\theta)p_w(\theta)^m} \tag{4.4}$$

### Resultados Técnicos Auxiliares

En esta sección se presentan los resultados técnicos preliminares que se usarán para demostrar el Teorema 4.1. EL análisis se refiere a un sistema especial de ecuaciones lineales

$$\mathbf{A}\mathbf{X} = \mathbf{Y},$$

donde, para ciertos números  $\mathbf{a}, b \in \mathbf{R}$ , la matriz  $\mathbf{A}$  de orden  $n \times n$  está dada por

$$\mathbf{A}_{n \times n} = \mathbf{A} = \begin{bmatrix} 1-a & -b & 0 & 0 & \dots & 0 & 0 \\ -a & 1 & -b & 0 & \dots & 0 & 0 \\ -a & 0 & 1 & -b & \dots & 0 & 0 \\ -a & 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -a & 0 & 0 & 0 & \dots & 1 & -b \\ -a & 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix} \tag{5.1}$$

y el vector  $\mathbf{Y}$  tiene todas sus componentes iguales a uno, esto es,

$$\mathbf{Y} = \mathbf{I} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \tag{5.2}$$

Por conveniencia notacional, las filas y columnas de una matriz se numeran empezando desde cero, con una convención similar para vectores. El resultado principal de esta sección es el siguiente:

**Teorema 5.1.** (i) Para la matriz  $\mathbf{A}_{n \times n}$ , definida en (5.1) se tiene que

$$\det \mathbf{A} = \begin{cases} 1 - na; & \text{si } b = 1 \\ \frac{1 - b - a(1 - b^n)}{1 - b}; & \text{si } b \neq 1 \end{cases}$$

(ii) Si  $a$  y  $b$  son dos números no negativos que satisfacen  $a + b \leq 1$ , entonces  $\det A \neq 0$ .

Más aún, en este caso,

(iii) La única solución al sistema  $A\mathbf{x} = \mathbf{1}$  es el vector  $\mathbf{x}$  cuyas componentes son

$$x_k = \frac{1 - b^{n-k}}{1 - b - a(1 - b^n)}, \quad k = 0, 1, 2, \dots, n-1$$

**Demostración.** (i) Para calcular el determinante de la matriz  $A_{n,n}$ , se utilizará el desarrollo a través de la última fila; ésta es una elección conveniente, entre otras razones, debido a que sólo dos elementos en dicha fila son no nulos, a saber, el primero (con valor  $-a$ ) y el último (cuyo valor es  $1$ ). Realizando la expansión, el coeficiente de  $-a$  es  $(-1)^{n+1}$  multiplicado por el determinante de la matriz que se obtiene eliminando la primera columna y última fila de  $A$ , mientras que el coeficiente de  $1$  es el determinante de la matriz que resulta al eliminar la última fila y la última columna de  $A$  multiplicado por  $(-1)^{n+n} = 1$ :

$$\det A_{n,n} = (-1)^{n+1} (-a) \cdot \det \begin{bmatrix} -b & 0 & 0 & \dots & 0 & 0 \\ 1 & -b & 0 & \dots & 0 & 0 \\ 0 & 1 & -b & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -b & 0 \\ 0 & 0 & 0 & \dots & 1 & -b \end{bmatrix} + (1) \times \det \begin{bmatrix} 1-a & -b & 0 & \dots & 0 & 0 \\ -a & 1 & -b & \dots & 0 & 0 \\ -a & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -a & 0 & 0 & \dots & 1 & -b \\ -a & 0 & 0 & \dots & 0 & 1 \end{bmatrix}$$

Las dos matrices cuyo determinante aparece en el lado derecho de esta igualdad son de orden  $(n-1) \times (n-1)$ . La primera es triangular superior con sus elementos diagonales iguales a  $-b$ , por lo que su determinante es  $(-b)^{n-1}$ , mientras que la segunda matriz es no es otra cosa que la versión de  $A$  en (5.1) con orden  $(n-1) \times (n-1)$ . Luego,

$$\det A_{n,n} = (-1)^{n+1} (-a) \times (-b)^{n-1} + (1) \times \det A_{(n-1) \times (n-1)} = -ab^{n-1} + \det A_{(n-1) \times (n-1)} \tag{5.3}$$

Esta ecuación permite expresar el determinante de una matriz  $A$  en términos de otra matriz del mismo tipo, pero de orden menor; una fórmula de esta clase se denomina recursiva. Aplicando (5.3), pero ahora empezando con una matriz de orden  $(n-1) \times (n-1)$  se obtiene que  $\det A_{(n-1) \times (n-1)} = -ab^{n-2} + \det A_{(n-2) \times (n-2)}$ , y reemplazando esta expresión en (5.3) se obtiene

$$\det A_{n,n} = -ab^n + \det A_{(n-2) \times (n-2)}$$

El mismo procedimiento se puede aplicar repetidamente para llegar a expresar  $\det \mathbf{A}_{n \times n}$  en términos de un determinante que pueda ser fácilmente calculable, digamos hasta llegar a la expresión que involucra  $\det \mathbf{A}_{2 \times 2}$ . Haciendo esta tarea se obtiene

Observe ahora que  $\det \mathbf{A}_{2 \times 2} = \det \begin{bmatrix} 1-a & -b \\ -a & 1 \end{bmatrix} = 1-a-ab$ , igualdad que al combinarse con la anterior ecuación implica

$$\begin{aligned} \det \mathbf{A}_{n \times n} &= -ab^{n-1} - ab^{n-2} - \dots - ab^2 + [1-a-ab] \\ &= 1-a[b^{n-1} + b^{n-2} + \dots + b^2 + b + 1] \end{aligned} \tag{5.4}$$

A continuación, note que, para  $b=1$  la suma entre corchetes es

$$[b^{n-1} + b^{n-2} + \dots + b^2 + b + 1] = [1+1+\dots+1+1] = n,$$

de manera que

$$\det \mathbf{A}_{n \times n} \Big|_{b=1} = 1 - a n$$

mientras que para  $b \neq 1$ , se aplica la fórmula para la adición de términos que cambian geoméricamente, obteniendo

$$[b^{n-1} + b^{n-2} + \dots + b^2 + b + 1] = \frac{1-b^n}{1-b}, \quad b \neq 1,$$

y entonces (5.4) implica que

$$\det \mathbf{A}_{n \times n} = 1 - a \frac{1-b^n}{1-b} = \frac{1-b-a(1-b^n)}{1-b}, \quad \text{si } b \neq 1.$$

estableciendo la conclusión deseada.

(ii) Suponga que  $a$  y  $b$  son dos números positivos con  $a+b \leq 1$ . En este caso, necesariamente se tiene que  $b < 1$  y entonces la parte (i) implica que

$$\det \mathbf{A}_{n \times n} = 1 - a \frac{1-b^n}{1-b} = \frac{1-b-a(1-b^n)}{1-b}$$

note ahora que  $a > a(1-b^n)$  y entonces  $1-b \geq b+a \geq b+a(1-b^n)$ , de donde se desprende que  $1-b-a(1-b^n) > 0$ , de manera que el numerador en el cociente para  $\det \mathbf{A}_{n \times n}$  es no nulo, con lo cual  $\det \mathbf{A}_{n \times n} \neq 0$ .

(iii) El hecho de que el vector  $X$  determinado por

$$x_k = \frac{1-b^{n-k}}{1-b-a(1-b^n)}, \quad k=0,1,2,\dots,n-1,$$

La solución de  $\mathbf{A}X = \mathbf{1}$  puede verificarse directamente, esto es, sustituyendo los valores postulados de las componentes de  $X$  en las ecuaciones del sistema  $\mathbf{A}X = \mathbf{1}$ . Debido a que  $\det \mathbf{A} \neq 0$ , ésta es la única solución del sistema (Hoffman y Kunze (1973), Grossman (1983), o Harville (1997)).

Antes de proceder a demostrar el Teorema 4.1, observe que debido a la igualdad  $p_{1,1}(i) + p_{2,1}(i) + p_{3,1}(i) = 1$  la ecuación (4.1) es equivalente a

(ii) Si  $a$  y  $b$  son dos números no negativos que satisfacen  $a + b \leq 1$ , entonces  $\det A \neq 0$ .

Más aún, en este caso.

(iii) La única solución al sistema  $AX = 1$  es el vector  $X$  cuyas componentes son

$$x_k = \frac{1 - b^{n-k}}{1 - b - a(1 - b^n)}, \quad k = 0, 1, 2, \dots, n-1$$

**Demostración.** (i) Para calcular el determinante de la matriz  $A_{n,n}$  se utilizará el desarrollo a través de la última fila; ésta es una elección conveniente, entre otras razones, debido a que sólo dos elementos en dicha fila son no nulos, a saber, el primero (con valor  $-a$ ) y el último (cuyo valor es  $1$ ). Realizando la expansión, el coeficiente de  $-a$  es  $(-1)^{n+1}$  multiplicado por el determinante de la matriz que se obtiene eliminando la primera columna y última fila de  $A$ , mientras que el coeficiente de  $1$  es el determinante de la matriz que resulta al eliminar la última fila y la última columna de  $A$  multiplicado por  $(-1)^{2n} = 1$ :

$$\det A_{n,n} = (-1)^{n+1} (-a) \cdot \det \begin{bmatrix} -b & 0 & 0 & \dots & 0 & 0 \\ 1 & -b & 0 & \dots & 0 & 0 \\ 0 & 1 & -b & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -b & 0 \\ 0 & 0 & 0 & \dots & 1 & -b \end{bmatrix} + (1) \times \det \begin{bmatrix} 1-a & -b & 0 & \dots & 0 & 0 \\ -a & 1 & -b & \dots & 0 & 0 \\ -a & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -a & 0 & 0 & \dots & 1 & -b \\ -a & 0 & 0 & \dots & 0 & 1 \end{bmatrix}$$

Las dos matrices cuyo determinante aparece en el lado derecho de esta igualdad son de orden  $(n-1) \times (n-1)$ . La primera es triangular superior con sus elementos diagonales iguales a  $-b$ , por lo que su determinante es  $(-b)^{n-1}$ , mientras que la segunda matriz es no es otra cosa que la versión de  $A$  en (5.1) con orden  $(n-1) \times (n-1)$ . Luego,

$$\det A_n = (-1)^{n+1} (-a) \times (-b)^{n-1} + (1) \times \det A_{(n-1) \times (n-1)} = -ab^{n-1} + \det A_{(n-1) \times (n-1)} \tag{5.3}$$

Esta ecuación permite expresar el determinante de una matriz  $A$  en términos de otra matriz del mismo tipo, pero de orden menor; una fórmula de esta clase se denomina recursiva. Aplicando (5.3), pero ahora empezando con una matriz de orden  $(n-1) \times (n-1)$  se obtiene que  $\det A_{(n-1) \times (n-1)} = -ab^{n-2} + \det A_{(n-2) \times (n-2)}$ , y reemplazando esta expresión en (5.3) se obtiene

$$\det A_{n,n} = -ab^{n-1} - ab^{n-2} + \det A_{(n-2) \times (n-2)}$$

además,  $n = m$ . Observe ahora que en el contexto de la norma  $RP_{1-m}$ , las probabilidades  $p_{i,j}(\theta)$  y  $p_{j,i}(\theta)$  son claramente no negativas y satisfacen

$$p_{i,j}(\theta) + p_{j,i}(\theta) \leq 1,$$

de manera que las condiciones del Teorema 5.1 (ii) se verifican, implicando que el sistema (5.7) tiene la única solución dada por

$$\begin{aligned} E_{\theta}[N|C_k] &= x_k \\ &= \frac{1 - b^{m-k}}{1 - b - a(1 - b^n)} \\ &= \frac{1 - p_{w^*}(\theta)^{m-k}}{1 - p_{w^*}(\theta) - p_{s^*}(\theta)(1 - p_{w^*}(\theta)^m)} \end{aligned}$$

en particular, la longitud promedio de una corrida es

$$LPC(\theta) = E_{\theta}[N|C_0] = \frac{1 - p_{w^*}(\theta)^m}{1 - p_{w^*}(\theta) - p_{s^*}(\theta)(1 - p_{w^*}(\theta)^m)}$$

### Conclusiones

En este trabajo se ha formulado una norma de paro general derivada de un gráfico de control y su correspondiente índice de funcionamiento se ha determinado sin requerir la distribución probabilística de la longitud de una corrida de datos. Los argumentos presentados fueron motivados por técnicas empleadas en el estudio de problemas de programación dinámica, y se basan en argumentos de esperanza condicional, los cuales permiten establecer un sistema de ecuaciones lineales en cuya solución se incluye al indicador del desempeño de la regla general de paro; de esta forma, el enfoque de programación dinámica permite abordar el problema de calcular el índice de funcionamiento a través de una perspectiva algebraica.

### Literatura Citada

1. Cavazos-Cadena, R. (1991), A counter example on the optimality equation in Markov decision chains with the average cost criterion, *Systems & Control Letters*, 16, 387-392.
2. Cavazos-Cadena, R. (1996), Value iteration in a class of communicating Markov decision chains with the average cost criterion, *SIAM Journal on Control and Optimization*, 36, 1848-1873.
3. Duncan, A. J. (1989), *Control de Calidad y Estadística Industrial, Alfaomega*, México, D.F.
4. Godfrey, A. B. y Mundel A. (1984), Guide for selection of an acceptance sampling plan, *Journal of Quality Technology*, 16,50-55.
5. Graybill, F. A. (1985), *Theory and Application of the Linear Model, Wadsworth*, Belmont, CA
6. Grossman, S. I. (1983), *Algebra Lineal, Grupo Editorial Iberoamérica*, México D. F.
7. Harville D. A. (1997), *Matrix Algebra from a Statistician Perspective, Spaingner-Verlag*, New York.
8. Hofhan, K. y R. Kunze (1973), *Algebra Lineal, Prentice Hall Hispanoamericana*, México, D.F.
9. John, P. W. M. (1990), *Statistical Methods in Engineering and Quality Assurance, Jeley*, New York.
10. Montgomery, D. C. (1985), *Statistical Quality Control, Wiley*, New York.
11. Puterman M. (1994), *Markov Decision Processes, Wiley*, New York.
12. Shewhart, W. A. (1931), *The Economic Control of Quality of Manufactured Product, Van Nostrand*, New York.
13. Sennoti L. 1, (1995), *Another set of conditions for average optimality in Markov control processes, Systems & Control Letters*, 24, 147-151.
14. Sennoti L. 1, (1998), *Stochastic Dynamic Programming and the Control of Queuing Systems, Wiley*, New York.

# Una Aplicación de los Algoritmos Genéticos en la Discriminación

**Aurora Montano Rivas**

**Mario Cantú Sifuentes**

*Departamento de Estadística y Cálculo, Universidad Autónoma Agraria Antonio Narro  
mcansi@uaaan.mx*

**Nidia Hernández Pérez**

*In memoriam*

---

**Abstract.** *An application of genetic algorithms in discrimination. In this work a technique based on genetic algorithms is presented. Such a technique produces better discriminant functions parameters estimators, in the sense that the clasiffication errors are reduced. The technique is applied to real data, and the results are compared with the ones produced by means of the clasical Fisher discriminant analysis.*

**Key Words:** Discrete Optimization, Heuristic Algorithms.

**Resumen.** En este trabajo se presenta una técnica basada en algoritmos genéticos la cual produce mejores estimadores de los parámetros de funciones discriminantes que los encontrados por métodos tradicionales; en el sentido de que con la técnica propuesta se disminuyen los errores de clasificación. Se aplica la técnica a datos reales y los resultados se comparan con los que produce el análisis discriminante clásico de Fisher.

**Palabras clave:** Optimización discreta, Algoritmos Heurísticos.

---

## Introducción

El análisis discriminante forma parte del conjunto de técnicas estadísticas diseñadas para resolver el problema de clasificación. Se caracteriza por estudiar la relación entre una variable categórica dependiente (el grupo de clasificación) y un conjunto de variables reales (posiblemente vectoriales) independientes. Estas describen a cada uno de los individuos a clasificar. A través de esta técnica, se construyen reglas de decisión que permiten discriminar o separar grupos mediante funciones de las variables observadas, minimizando la probabilidad de clasificación errónea, o bien maximizandola si la clasificación es correcta. Las reglas más usuales se construyen a partir de un modelo probabilístico, y de muestras de entrenamiento.

Por otra parte, los algoritmos genéticos son métodos adaptativos heurísticos que pueden usarse para resolver problemas de búsqueda y de optimización sobre el espacio de soluciones posibles (soluciones factibles). Esta técnica

está inspirada en los mecanismos de la genética y de la selección natural. Dada una función objetivo, los algoritmos genéticos construyen el espacio de soluciones usando operadores genéticos, entre ellos: selección, cruzamiento, y mutación. En cada etapa de la búsqueda, sobreviven las soluciones subóptimas.

En este trabajo se presenta una aplicación de los algoritmos genéticos a las funciones discriminantes que conforman la población inicial, las cuales se generan realizando el análisis discriminante de cada muestra obtenida por la técnica de remuestreo que se aplicó a una muestra de entrenamiento. Posteriormente, las funciones se evalúan para seleccionar las mejores y, a través de la aplicación de operadores genéticos, llegar a obtener una o un grupo de funciones que discriminen mejor que la proporcionada por el método de Fisher.

Para comprender adecuadamente el desarrollo de la aplicación de la técnica propuesta, se consideró conveniente presentar una breve descripción de los

métodos empleados: el análisis discriminante, los algoritmos genéticos, los métodos de selección y los operadores genéticos; también se describe la aplicación, seguida de las conclusiones y las referencias.

### El Análisis Discriminante

Bajo el supuesto de  $g$  grupos o poblaciones  $\Pi_i, i = 1, \dots, g$  mutuamente excluyentes, con probabilidades a priori asociadas a cada población como  $\alpha_i, \alpha_i > 0$  y  $\sum \alpha_i = 1$  y vector de variables aleatorias  $X = (X_1, \dots, X_p)$ , el objetivo que se persigue es obtener un conjunto de  $g(g-1)/2$  funciones lineales o seudonlineales del vector  $X$  tal, que maximice la separación entre los grupos y permita minimizar la probabilidad de clasificación errónea en cada grupo. Bajo los supuestos de que  $g = 2$  y de que la separación entre los grupos es lineal o pseudolineal, una regla de decisión para este caso da por:

La unidad  $u$  con vector de variables observadas  $x$  clasifica en  $\Pi_1$  si

$$a'x \leq 0$$

La unidad  $u$  con vector de variables observadas  $x$  clasifica en  $\Pi_2$  si

$$a'x > 0$$

siendo  $a \in R^{p+1}$  tal, que minimice la probabilidad de clasificación errónea, es decir,

$$a^* = \min_{a \in R^{p+1}} (\alpha_1 P(a'x \leq 0 | x \in \Pi_1) + \alpha_2 P(a'x > 0 | x \in \Pi_2))$$

Dada una muestra de entrenamiento  $X$  de tamaño  $N \cdot p$  con  $N = \sum n_j, j = 1, 2$  y la regla de clasificación anterior, se construye la regla de clasificación empírica a partir de la función discriminante estimada  $\hat{a}'(X)x$  al ser  $\hat{a}'(X)$  una estimación del vector de parámetros  $a$  tal, que minimice la estimación de la probabilidad de clasificación errónea condicionada a la muestra de entrenamiento, es decir, la selección del vector  $a$  debe cumplir:

$$(a^* | X) = \min_{a \in R^{p+1}} (a, \hat{P}(a'(X)x \leq 0 | x \in \Pi_1, X) + \alpha_2 \hat{P}(a'(X)x > 0 | x \in \Pi_2, X))$$

con

$$\hat{P}(a'(X)x > 0 | x \in \Pi_2, X) = \#(x \in \Pi_2 : a'(X)x > 0) / n_2$$

$$\hat{P}(a'(X)x \leq 0 | x \in \Pi_1, X) = \#(x \in \Pi_1 : a'(X)x \leq 0) / n_1$$

### Algoritmos Genéticos

Los algoritmos genéticos buscan el mejor resultado a través del uso del espacio de posibles soluciones del problema considerado.

Estos comienzan con una población inicial o población cero, que representa un conjunto de posibles soluciones generadas aleatoriamente. Cada solución del problema planteado representa un cromosoma. Generalmente este conjunto discreto de soluciones es de tamaño de 20 a 100 cromosomas.

Tomando en consideración el proceso evolutivo de Danvin, la estrategia de estos algoritmos es seleccionar, de la población inicial, los mejores cromosomas padres como candidatos a producir mejores hijos en la nueva generación. Se repite este proceso hasta localizar una solución óptima o subóptima.

### Selección de mejores cromosomas

La idea es escoger los mejores cromosomas o soluciones en una población que se va a reproducir para formar generaciones sucesivas; esta selección se realiza aleatoriamente, y se asigna una probabilidad  $P_j$  a cada elemento  $j$ , perteneciente a la población en cuestión y se toma como base el valor de ajuste. También se generan números aleatorios con distribución uniforme, uno para cada elemento de la población, y se comparan contra la probabilidad acumulada

$$c_i = \sum_{j=1}^L P_j$$

Entonces, la solución o cromosoma  $i$  se selecciona para integrar la nueva población, si

$$c_i < U(0,1) < c_{i+1}$$

**Operadores genéticos: cruzamiento y mutación**

Los operadores genéticos proporcionan los mecanismos de búsqueda básicos de los algoritmos genéticos; los cuales se usan para crear nuevas soluciones, basadas en el conjunto de las mejores soluciones escogidas previamente por algún método de selección.

Básicamente se usan dos operadores para valores reales, lo cuales son: cruzamiento y mutación. Operador cruzamiento: selecciona dos soluciones (cromosomas progenitores) y produce dos nuevas soluciones (nuevos cromosomas). Operador mutación: en este operador un gene se selecciona aleatoriamente de una cadena, el cual sufre una alteración y se produce una nueva solución única.

La aplicación de los operadores depende de la estructura de los cromosomas, si son discretos o continuos. Para mayor información puede verse en Michalewicz (1994) y Houck, et al., (1998). En el problema que se presenta se utilizaron los siguientes operadores: Cruzamiento aritmético: sean  $S_j$  y  $S_k$  dos soluciones o cromosomas, entonces al seleccionar aleatoriamente un valor  $r$  de una variable aleatoria con distribución  $U(0,1)$ , se logra la generación de dos nuevas soluciones o nuevos cromosomas al considerar las combinaciones convexas entre los dos, es decir:

$$S_m' = rS_m + (1-r)S_k$$

$$S_k' = rS_k + (1-r)S_m$$

De esta forma se generan dos nuevas soluciones. Mutación uniforme: sea  $S(i)$  la matriz de tamaño  $s \times p$  cuyas filas representan  $s$  posibles soluciones del problema en la iteración  $i$  con  $p$  componentes, y

$$a_j = \min_j S_{j,}$$

$$h_j = \max_j S_{j,}$$

donde  $j = 1, \dots, p$ .

es decir, el límite inferior y superior del conjunto de cada columna de la matriz  $S(i)$  del problema. Seleccionando aleatoriamente el valor  $k$  entre 1 y  $p$  y un valor de la variable aleatoria  $F'(a_k, h_k)$  con distribución uniforme en el intervalo  $(a_k, h_k)$ , se realiza la siguiente mutación:

$$S_{j,} = U(a_k, h_k), \text{ si } j = k$$

$$j \neq k$$

**Aplicación**

Para ilustrar el comportamiento de esta técnica, se usó la base de datos de IRIS dada por Fisher (1936). SE consideran solamente con especies (Versicolor y Virginica), con un total de 50 casos y cuatro variables:

- $X_1$  = longitud del sépalo
- $X_2$  = ancho del sépalo
- $X_3$  = longitud del pétalo
- $X_4$  = ancho del pétalo.

Los cálculos del procedimiento propuesto anteriormente, se realizaron con el apoyo de los paquetes SPLUS 2000 y MATLAB 12 .

La solución obtenida bajo los supuestos de normalidad y de homogenidad de las matrices de covarianza de las dos especies, y considerando la muestra de entrenamiento de tamaño 100, se obtuvo como vector  $(\mu^T, \Sigma)$  de coeficientes de la función discriminante  $(3,5563,55786,-6.97013,-12386,16.663)$ , con una estimación del error de clasificación de 0.03.

Después de aplicar las técnicas de remuestreo a la muestra original, se generaron 12 muestras aleatorias, a las cuales se les aplicó la discriminación bajo los mismos supuestos. Las soluciones dadas del vector  $(\mu^T, \Sigma)$  para cada muestra aparecen en el Cuadro 1, que muestran los cromosomas progenitores o población inicial. La última columna del cuadro indica el error estimado, resultante de reclasificar la muestra de entrenamiento original con las 12 soluciones resultantes.

Asignando una probabilidad frecuencia 1 a cada error y al generar 12 números aleatorios con distribución  $U(0,1)$ , ordenados de menor a mayor, se procede a utilizar la técnica de selección. Los Cuadros 2 y 3 muestran los resultados de la primera selección. A continuación se realizó el primer cruce aritmético (Cuadro 4) y se obtuvieron las nuevas combinaciones o nuevas soluciones.

Para realizar la primera mutación uniforme, se seleccionó un número aleatorio entre 1 y 5, en este caso resultó  $j = 4$  y se generó un número aleatorio (gene) entre los límites superior e inferior del vector de soluciones para la columna 4. El valor obtenido fue 13.2639, el cual se sustituyó en la columna correspondiente para cada uno de los cromosomas. En el Cuadro 5 se dan los resultados obtenidos.

Los Cuadros 6, 7 y 8 muestran los resultados del segundo cruce. Para realizar la segunda mutación se seleccionó un número aleatorio entre 1 y 5, sin considerar el 4, que ya corresponde a la columna mutada

anteriormente. En este caso, la nueva columna a mutar es  $j = 5$  y el valor aleatorio elegido entre 15.0474 y 15.2138 es 15.1394. En el Cuadro 9 se dan los resultados.

Al repetir el algoritmo anterior, se consideraron cinco iteraciones que se muestran en los Cuadros 10, 11, 12, 13, 14, 15, 16, 17 y 18 que se presentan en el Anexo.

Puede observarse como la estimación del error de clasificación va estabilizándose a 0.02. Una solución subóptima del problema de discriminación original es dada por la función discriminante lineal:

$$6.3981X_1 + 5.0345X_2 - 9.7445X_3 - 13.2639X_4 + 15.1394$$

con una estimación del error de clasificación de 0.02.

### Conclusiones

En este trabajo se ha mostrado que la técnica de algoritmos genéticos puede usarse para refinar los métodos clásicos de discriminación. Esta afirmación es soportada por el hecho de que el método de clasificación de Fisher aplicado al conjunto de datos de IRIS, tiene un error de clasificación de 0.03; al aplicar algoritmos genéticos el método se mejora al obtenerse un error de clasificación de 0.02 o menor.

### Literatura Citada

1. Barbro, B., Teija, L. y Kaisa, S. 1996. Bankruptcy Predictors Using Discriminant Analysis, Logit Analysis, and Genetic Algorithms. Turku, Centre for Computer Science, Finland. Technical Report No.40.
2. Fisher R. A. 1936. The Use of Multiple Measurements in Taxonomic Problems. Annual Eugenics 7:179-188.
3. Houck, C. R. Joines J.A., y Kay M. G. 1990. A genetic Algorithm for Function Optimization: A Matlab Implementation, NCSU-IE, Technical Report 95-09.
4. *MATLAB R. 12*. The Math Works Company, Natick, Massachusetts.
5. Michalewicz, Z 1994. Genetic Algorithms + Data Structures = Evolution Program. Series Springer-Verlag, New York, USA.
6. Mitsuru O. 2000. New Method Discriminant Analysis Utilizing Genetic Algorithm. Fujitsu Limited 4-1-1, Kamikodanaka, Nakahara-ku, Kawasaki, 211-8588.
7. *S-PLUS 2000*, Data analysis Products Division, MathSoft, Seattle, WA.

## ANEXO

**Cuadro 1.** Cromosomas Progenitores

Sepal	Sepaw	Petal	Petaw	Constant	Error
4.5665	7.2112	-8.3568	-14.0763	15.4254	.04
0.10279	7.49824	-3.5749	-14.9522	20.8127	.06
1.8046	6.21916	-4.3873	-16.1442	19.7121	.02
2.7065	6.2569	-6.1654	-12.889	17.7004	.038
4.5665	7.2112	-8.3568	-14.0763	15.4254	.04
3.556	5.5786	-6.97013	-12.386	16.663	.03
2.6769	8.234	-7.45	-16.3613	24.5204	.03
3.7958	8.027	-6.823	-16.5313	14.639	.046
4.7955	4.2296	-7.2728	-14.0795	16.89312	.02
1.8046	6.2192	-4.38726	-16.1442	19.7112	.02
6.6358	5.4506	-10.48465	-12.41914	14.9495	.01
6.3973	4.13926	-9.04369	-13.71768	15.15608	.00

**Cuadro 2.** Primera Selección de Cromosomas Progenitores

Error	Prob.Acum.	$(c_{i-1} - 1) \cdot U(0,1) \cdot c_i$	
.04	.083	(0 < .01 < .083)	Si
.06	.166	(.083 < .23 < .166)	No
.02	.249	(.166 < .44 < .249)	No
.038	.332	(.249 < .45 < .332)	No
.04	.415	(.332 < .48 < .415)	No
.03	.498	(.415 < .60 < .498)	No
.03	.581	(.498 < .61 < .581)	No
.046	.664	(.581 < .76 < .664)	No
.02	.747	(.664 < .79 < .747)	No
.02	.83	(.747 < .82 < .83)	Si
.01	.913	(.83 < .89 < .913)	Si
.00	1	(.913 < .95 < 1)	Si

**Cuadro 3.** Combinaciones Seleccionadas

Sepal	Sepaw	Petal	Petaw	Constant	Error
4.5665	7.2112	-8.3568	-14.0763	15.4254	.04
1.8046	6.2192	-4.38726	-16.1442	19.7112	.02
6.6358	5.4506	-10.48465	-12.41914	14.9495	.01
6.3973	4.13926	-9.04369	-13.71768	15.15608	.00

**Cuadro 4.** Resulta del Primer Cruce Aritmético

Sepal	Sepaw	Petal	Petaw	Constant	Error
3.07507	6.67555	-6.213248	-15.192968	17.740184	.02
3.29602	6.7548	-6.5308	-15.0275	17.397266	.02
5.3801	6.5189	-9.1935	-13.4247	15.2383	.03
5.8222	6.1429	-9.648	-13.0707	15.1366	.03
5.5075	5.6322	-8.7099	-13.8919	15.2870	.03
5.4563	5.7182	-8.6906	-13.902	15.2945	.03
2.3298	6.1357	-5.05	-15.7393	19.1944	.02
6.1106	5.5341	-9.8219	-12.8241	15.4672	.03
2.8972	5.7244	-5.4950	-15.5669	18.6282	.03
5.3047	4.6341	-7.9359	-14.2949	16.24	.03
6.5062	4.7379	-9.7015	-13.1249	15.0618	.03
6.5269	4.8520	-9.8269	-13.0119	15.0438	.03

**Cuadro 5.** Primera Mutuación Uniforme

Sepal	Sepaw	Petal	Petaw	Constant	Error
3.07507	6.67555	-6.213248	-13.2639	17.740184	.12
3.29602	6.7548	-6.5308	-13.2639	17.397266	.08
5.3801	6.5189	-9.1935	-13.2639	15.2383	.03
5.8222	6.1429	-9.648	-13.2639	15.1366	.02
5.5075	5.6322	-8.7099	-13.2639	15.2870	.03
5.4563	5.7182	-8.6906	-13.2639	15.2945	.03
2.3298	6.1357	-5.05	-13.2639	19.1944	.15
6.1106	5.5341	-9.8219	-13.2639	15.4672	.03
2.8972	5.7244	-5.4950	-13.2639	18.6282	.14
5.3047	4.6341	-7.9359	-13.2639	16.24	.05
6.5062	4.7379	-9.7015	-13.2639	15.0618	.02
6.5269	4.8520	-9.8269	-13.2639	15.0438	.02

**Cuadro 6.** Primera Selección

Error	Prob.Acum.	$(\epsilon_i - 1) \cdot U(0,1) - \epsilon_i$	
.12	.083	(0 < .151 < .083)	No
.08	.166	(.083 < .235 < .166)	No
.03	.249	(.166 < .240 < .249)	Si
.02	.332	(.249 < .334 < .332)	No
.03	.415	(.332 < .591 < .415)	No
.03	.498	(.415 < .718 < .498)	No
.15	.581	(.498 < .738 < .581)	No
.03	.664	(.581 < .792 < .664)	No
.14	.747	(.664 < .875 < .747)	No
.05	.83	(.747 < .894 < .83)	No
.02	.913	(.83 < .897 < .913)	Si
.02	1	(.913 < .939 < 1)	Si

**Cuadro 7.** Combinaciones Seleccionadas

Sepal	Sepaw	Petal	Petaw	Constant	Error
5.3801	6.5189	-9.1935	-13.2639	15.2383	.03
6.5062	4.7379	-9.7015	-13.2639	15.0618	.02
6.5269	4.8520	-9.8269	-13.2639	15.0438	.02

**Cuadro 8.** Segundo Cruce Aritmético

Sepal	Sepaw	Petal	Petaw	Constant	Error
6.3498	4.9853	-9.6309	-13.2639	15.0863	.03
5.5365	6.2715	-9.2641	-13.2639	15.2138	.03
6.2943	5.19	-9.6984	-13.2639	15.0832	.02
5.6127	6.1809	-9.3220	-13.2639	15.1989	.03
6.5228	4.8293	-9.8020	-13.2639	15.0474	.02
6.5103	4.7606	-9.7264	-13.2639	15.0582	.02

**Cuadro 9.** Segunda Mutación

Sepal	Sepaw	Petal	Petaw	Constant	Error
6.3498	4.9853	-9.6309	-13.2639	15.1394	.03
5.5365	6.2715	-9.2641	-13.2639	15.1394	.03
6.2943	5.19	-9.6984	-13.2639	15.1394	.02
5.6127	6.1809	-9.3220	-13.2639	15.1394	.03
6.5228	4.8293	-9.8020	-13.2639	15.1394	.02
6.5103	4.7606	-9.7264	-13.2639	15.1394	.02

**Cuadro 10.** Tercera Selección

Error	Prob.Acum.	$(\epsilon_i - 1) \cdot U(0,1) - \epsilon_i$	
.03	.167	(0 < .243 < .167)	No
.03	.333	(.167 < .361 < .333)	No
.02	.5	(.333 < .412 < .5)	Si
.03	.666	(.5 < .434 < .666)	No
.02	.833	(.666 < .712 < .833)	Si
.02	1	(.833 < .982 < 1)	Si

**Cuadro 11.** Combinaciones Seleccionadas

Sepal	Sepaw	Petal	Petaw	Constant	Error
6.2943	5.19	-9.6984	-13.2639	15.1394	.02
6.5228	4.8293	-9.8020	-13.2639	15.1394	.02
6.5103	4.7606	-9.7264	-13.2639	15.1394	.02

**Cuadro 12.** Tercer Cruce Aritmético

Sepal	Sepaw	Petal	Petaw	Constant	Error
6.47	4.9127	-9.7781	-13.2639	15.1394	.02
6.3471	5.1066	-9.7223	-13.2639	15.1394	.02
6.4722	4.8363	-9.7215	-13.2639	15.1394	.02
6.3324	5.1143	-9.7033	-13.2639	15.1394	.02
6.5154	4.7885	-9.7571	-13.2639	15.1394	.02
6.5177	4.8014	-9.7713	-13.2639	15.1394	.02

**Cuadro 13.** Tercera Mutación

Sepal	Sepaw	Petal	Petaw	Constant	Error
6.47	4.9127	-9.7445	-13.2639	15.1394	.03
6.3471	5.1066	-9.7445	-13.2639	15.1394	.02
6.4722	4.8363	-9.7445	-13.2639	15.1394	.02
6.3324	5.1143	-9.7445	-13.2639	15.1394	.02
6.5154	4.7885	-9.7445	-13.2639	15.1394	.03
6.5177	4.8014	-9.7445	-13.2639	15.1394	.03

**Cuadro 14.** Cuarta Selección

Error	Prob.Acum.	$(\epsilon_i - 1) \cdot U(0,1) \cdot \epsilon_i$	
.03	.167	(0 < .073 < .167)	Si
.02	.333	(.167 < .332 < .333)	Si
.02	.5	(.333 < .399 < .5)	Si
.02	.666	(.5 < .538 < .666)	Si
.03	.833	(.666 < .656 < .833)	No
.03	1	(.833 < .683 < 1)	No

**Cuadro 15.** Cuarto Cruce con las combinaciones seleccionadas

Sepal	Sepaw	Petal	Petaw	Constant	Error
6.3767	5.0599	-9.7445	-13.2639	15.1394	.02
6.4404	4.9594	-9.7445	-13.2639	15.1394	.03
6.4716	4.8585	-9.7445	-13.2639	15.1394	.02
6.4706	4.8905	-9.7445	-13.2639	15.1394	.03
6.369	5.0607	-9.7445	-13.2639	15.1394	.02
6.4334	4.9663	-9.7445	-13.2639	15.1394	.02
6.348	5.1047	-9.7445	-13.2639	15.1394	.02
6.4713	4.8382	-9.7445	-13.2639	15.1394	.02
6.3399	5.1104	-9.7445	-13.2639	15.1394	.02
6.3396	5.1105	-9.7445	-13.2639	15.1394	.02
6.4655	4.8496	-9.7445	-13.2639	15.1394	.02
6.3391	5.1010	-9.7445	-13.2639	15.1394	.02

**Cuadro 16.** Cuarta Mutación

Sepal	Sepaw	Petal	Petaw	Constant	Error
6.3767	5.0345	-9.7445	-13.2639	15.1394	.02
6.4404	5.0345	-9.7445	-13.2639	15.1394	.03
6.4716	5.0345	-9.7445	-13.2639	15.1394	.03
6.4706	5.0345	-9.7445	-13.2639	15.1394	.03
6.369	5.0345	-9.7445	-13.2639	15.1394	.02
6.4334	5.0345	-9.7445	-13.2639	15.1394	.03
6.348	5.0345	-9.7445	-13.2639	15.1394	.02
6.4713	5.0345	-9.7445	-13.2639	15.1394	.03
6.3399	5.0345	-9.7445	-13.2639	15.1394	.02
6.3396	5.0345	-9.7445	-13.2639	15.1394	.02
6.4655	5.0345	-9.7445	-13.2639	15.1394	.03
6.3391	5.0345	-9.7445	-13.2639	15.1394	.02

**Cuadro 17.** Quinta Selección

Error	Prob.Acum.	$(\epsilon_i - 1) \cdot U(0,1) \cdot \epsilon_i$	
.02	.083	(0 < .073 < .083)	Si
.03	.166	(.083 < .332 < .166)	No
.03	.249	(.166 < .399 < .249)	Si
.03	.332	(.249 < .538 < .332)	No
.02	.415	(.332 < .656 < .415)	No
.03	.498	(.415 < .683 < .498)	No
.02	.581	(.498 < .073 < .581)	No
.03	.664	(.581 < .332 < .664)	No
.02	.747	(.664 < .399 < .747)	No
.02	.83	(.747 < .538 < .83)	No
.03	.913	(.83 < .656 < .913)	No
.02	1	(.913 < .683 < 1)	Si

**Cuadro 18.** Quinto Cruce

Sepal	Sepaw	Petal	Petaw	Constant	Error
6.4331	5.0345	-9.7445	-13.2639	15.1394	.03
6.4152	5.0345	-9.7445	-13.2639	15.1394	.03
6.3736	5.0345	-9.7445	-13.2639	15.1394	.02
6.3422	5.0345	-9.7445	-13.2639	15.1394	.02
6.4575	5.0345	-9.7445	-13.2639	15.1394	.03
6.3532	5.0345	-9.7445	-13.2639	15.1394	.02





Universidad Autónoma Agraria Antonio Narro  
Buenavista Saltillo, Coah. México. CP 25315  
[www.uaaan.mx](http://www.uaaan.mx)

## Oferta Educativa

### Agronomía

- Ingeniero Agrónomo
- Ingeniero Agrónomo en Desarrollo Rural
- Ingeniero Agrónomo en Horticultura
- Ingeniero Agrónomo en Producción
- Ingeniero Agrónomo Parasitólogo
- Ingeniero Agrónomo en Agrobiología
- Maestría en Horticultura
- Maestría Profesional en Tecnología de Granos y Semillas \*\*
- Maestría en Parasitología Agrícola \*\*
- Doctorado en Parasitología Agrícola
- Maestría en Fitomejoramiento \*\*
- Doctorado en Fitomejoramiento \*\*
- Maestría en Ciencias Agrarias
- Doctorado en Ciencias Agrarias

### Ingeniería

- Ingeniero en Ciencia y Tecnología de Alimentos
- Ingeniero Agrónomo en Irrigación
- Ingeniero Mecánico Agrícola \*
- Maestría en Estadística Experimental \*\*
- Maestría en Ingeniería de Sistemas Agrícolas
- Doctorado en Ingeniería de Sistemas Agrícolas

### Socioeconómicas

- Licenciado en Economía Agrícola y Agronegocios
- Ingeniero Agrónomo Administrador
- Maestría en Valuación Agropecuaria

### Ciencia Animal

- Médico Veterinario Zootecnista
- Ingeniero Agrónomo Zootecnista
- Maestría en Zootecnia \*\*
- Maestría Profesional en Administración de Sistemas de Zootecnia
- Doctorado en Zootecnia \*\*

### Ambiental

- Ingeniero Forestal
- Ingeniero en Agroecología
- Ingeniero en Procesos Ambientales
- Ingeniero Agrícola y Ambiental
- Maestría en Forestal

---

\*Programa Acreditado por COMEAA

\*\*Programa en PIFOP y Padrón de Excelencia de CONACYT

### INFORMES

**Licenciatura:**

[docencia@uaaan.mx](mailto:docencia@uaaan.mx)

Teléfono en Saltillo:

01-800-718-4330 (servicio sin costo)

Teléfono en Torreón:

01-800-718-3586 (servicio sin costo)

**Postgrado:**

[postgrado@uaaan.mx](mailto:postgrado@uaaan.mx)

Teléfonos (844) 411 03 33, 411 03 34



Universidad Autónoma Agraria Antonio Narro  
Buenavista, Saltillo, Coah. México

Tel. (844) 411-01-00 (Conmutador) - [www.uaaan.mx](http://www.uaaan.mx)