

01963

**UNIVERSIDAD NACIONAL AUTONOMA  
DE MEXICO**

---

---



FACULTAD DE PSICOLOGIA  
DIVISION DE ESTUDIOS DE POSGRADO

PROCEDIMIENTO BAYESIANO DE ESTIMACION DE  
CALIFICACIONES ANTE DIFERENTES TIPOS DE  
INCERTIDUMBRE

**T E S I S**

QUE PARA OBTENER EL GRADO DE

**MAESTRIA EN  
PSICOLOGIA EDUCATIVA**

P R E S E N T A :

**ADELINA ROSAS MERCADO**

DIRECTOR: DR. SERAFIN J. MERCADO DOMÉNECH

COMITÉ DE TESIS

DRA. DOLORES MERCADO CORONA

DR. JAVIER AGUILAR VILLALOBOS

DRA. SANDRA CASTAÑEDA FIGUEIRAS

MTRO. HAROLDO ELORZA PÉREZ TEJADA



FACULTAD  
DE PSICOLOGÍA

MEXICO, D.F.

2004



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

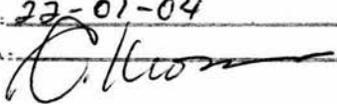
El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

**ESTA TESIS NO SALE  
DE LA BIBLIOTECA**

Autorizo a la Dirección General de Bibliotecas de la UNAM a difundir en formato electrónico e impreso el contenido de mi trabajo recepcional.

NOMBRE: ADELINA ROSAS  
MERCADO

FECHA: 22-01-04

FIRMA: 

*A Salvador  
mi compañero de toda la vida  
por su invaluable compañía y apoyo.*

*A mis hijos Adelina y Salvador*

*Al Dr. Serafin Mercado Doménech,  
un ser humano excepcional  
con mi más profundo agradecimiento  
y admiración.*

*Deseo expresar mi agradecimiento a la Facultad de Psicología y a la Universidad Nacional Autónoma de México con quienes tengo un compromiso para toda la vida. A mis maestros por dejar profunda huella en mí, en especial al Dr. Fausto Trejo Fuentes quien abrió la percepción y comprensión de mí misma y de mi entorno enseñándome a valorar a todas las personas y a luchar por mis ideales. Al Dr. Ely Rayek Zaga por distinguirme con su amistad y confianza. Al Dr. Luis Castro y a Martha por su amistad y por ayudarme a comprender la metodología de la investigación. A la Dra. Graciela Rodríguez Ortega por hacer patente en mí los valores a seguir en la práctica responsable de la psicología. Al doctor Gustavo Fernández y al Mtro. Haroldo Elorza por enseñarme metodología de la investigación y estadística, a éste último también, por darme la oportunidad de impartir con él un curso de estadística inferencial y al Dr. Samuel Jurado por su gentileza y apoyo.*

*Al Ing. Salvador Pérez Cárdenas por enseñarme a amar las matemáticas y a comprender la estadística bayesiana, por sus aportaciones, sugerencias y correcciones. Al Dr. Manuel González Hernández de la Escuela de Física y Matemáticas del Instituto Politécnico Nacional por revisar la congruencia matemática de este estudio y por haberme facilitado sus artículos. A la Lic. Adelina Pérez Rosas por sus sugerencias para mejorar el estilo y por su apoyo para realizar la presentación de esta tesis.*

*A los miembros del comité de sinodales: Dr. Javier Aguilar, Dra. Dolores Mercado, Dra. Sandra Castañeda Figueiras y Mtro. Haroldo Elorza por sus correcciones y útiles sugerencias para mejorar este trabajo.*

*Reitero mi agradecimiento al Dr. Serafín Mercado Doménech por su invaluable apoyo en todos los sentidos y por su paciencia.*

*Adelina Rosas Mercado*

*Ciudad Universitaria, diciembre de 2003.*

# Contenido

	Página
<b>Resumen</b>	
<b>Capítulo I</b>	
<b>Introducción</b>	1
<b>Antecedentes de la medición y evaluación psicológicas</b>	4
Medición y niveles de medición.	10
Evaluación y sus funciones.	14
Pruebas referidas a la norma.	17
Pruebas referidas a un criterio.	20
Asignación de calificaciones referida a un criterio <i>vs.</i> evaluación referida a la norma.	22
<b>Estudios bayesianos en evaluación educativa</b>	
Bayes: hacia una respuesta integral a la subjetividad-objetividad.	23
<b>Planteamiento del problema</b>	46
<b>Propósito y fundamentación</b>	47

## Capítulo II

<b>Método</b>	51
<b>Participantes</b>	52
<b>VARIABLES</b>	52
Dependientes	52
Independientes	53
<b>Diseño</b>	54
<b>Hipótesis</b>	55
Nula.	55
Alternativa.	55
<b>Procedimientos</b>	
<b>A priori objetiva</b>	56
Elección de una muestra aleatoria.	58
Determinación de las frecuencias de calificaciones iniciales y definitivas de la muestra.	61
Determinación de las probabilidades <i>a priori</i> objetivas.	63
Estimación de probabilidades de intersección de las calificaciones de la muestra.	64
Estimación de las probabilidades <i>a posteriori</i> con base en una <i>a priori</i> objetiva.	66
Estimación del valor esperado.	69

<i>A priori</i> no informativa	69
Estimación de probabilidades de intersección.	69
Cálculo de la probabilidad <i>a posteriori</i> .	70
Estimación del valor esperado	71
<i>A priori</i> subjetiva	72
Estimación de la probabilidad <i>a priori</i> .	72
Estimación de las probabilidades de intersección de la calificación inicial con la definitiva.	73
Estimación de probabilidades <i>a posteriori</i>	74
Estimación del valor esperado	75

## Capítulo III

### Resultados

Prueba de hipótesis: $\chi^2$ ji cuadrada.	76
Comparación gráfica de valores esperados.	81
Correlación entre estimadores de probabilidad y datos de la población.	83

## Capítulo IV

### Discusión

¿Cuál es el mejor procedimiento?	85
Posibles usos de los procedimientos bayesianos propuestos.	88
Limitaciones del estudio.	89
Sobre la objetividad y la subjetividad.	90
Líneas de investigación posibles.	93
Conclusiones.	94

<b>Nomenclatura</b>	96
<b>Glosario</b>	97
<b>Referencias bibliográficas</b>	106

## Tablas

	Página.
1. Calificaciones iniciales y definitivas de la población.	61
2. Frecuencias esperadas de la población.	62
3. Elementos de la muestra y frecuencias observadas según calificación inicial ( $c_1$ ) y calificación definitiva ( $c_d$ ).	63
4. Probabilidades de intersección de las calificaciones de la muestra.	66
5. Probabilidades <i>a posteriori</i> a partir de una <i>a priori</i> objetiva.	68
6. Probabilidades de intersección de las calificaciones para una <i>a priori</i> no informativa.	70
7. Probabilidades <i>a posteriori</i> para una <i>a priori</i> no informativa.	71
8. Probabilidades de intersección de la distribución subjetiva.	73
9. Probabilidades <i>a posteriori</i> a partir de una <i>a priori</i> subjetiva.	75
10. Frecuencias observadas y esperadas para los procedimientos de la calificación definitiva.	78

## Figuras

	Página
<b>Figura 1.</b> Representación esquemática del diseño de investigación de los efectos de diferentes tipos de <i>a priori</i> en la estimación bayesiana de calificaciones.	54
<b>Figura 2.</b> Probabilidad de obtener $NA_D$ dada la calificación inicial.	79
<b>Figura 3.</b> Probabilidad de obtener $S_D$ dada la calificación inicial.	80
<b>Figura 4.</b> Probabilidad de obtener $B_D MB_D$ dada la calificación inicial.	80
<b>Figura. 5.</b> Valor esperado para $NA$ inicial.	82
<b>Figura. 6.</b> Valor esperado para $S$ inicial.	82
<b>Figura. 7.</b> Valor esperado para calificación inicial $B$ o $MB$ .	83

## Resumen

Con base en el estudio de los efectos de tres tipos de probabilidades *a priori*: no informativa, subjetiva y objetiva, en la estimación bayesiana de calificaciones, se proponen dos procedimientos bayesianos de estimación de calificaciones o notas con fines de evaluación formativa. Las inferencias resultantes de la aplicación de los procedimientos bayesianos propuestos aplicados a una muestra de 185 alumnos de un curso de Fisicoquímica de la ESIQIE del IPN, fueron contrastadas con los datos reales de una población de 514 estudiantes. La prueba mostró resultados no significativos en el caso de la *a priori* objetiva y significativos al nivel del 0.01% en los otros casos, evidenciándose el alto grado de ajuste entre la estimación de probabilidad con base en la *a priori* objetiva y los datos reales de la población. Se obtuvieron los siguientes grados de exactitud para las diferentes probabilidades *a priori*: 1ro. Objetiva. 2do. Subjetiva. 3ero. No informativa. Los resultados de una prueba de correlación mostraron un valor de 0.98 para la *a priori* objetiva, único resultado significativo al nivel del 0.01%. Los procedimientos subjetivo y objetivo señalados permiten un amplio margen de acción en la toma de decisiones con fines de evaluación formativa, ante diversos tipos de información y tiempo disponibles, así como ante diversas funciones de pérdida.

Palabras clave: procedimientos bayesianos, evaluación formativa, estimación de calificaciones.

# Capítulo I

## Introducción

Este trabajo propone y compara tres procedimientos estadísticos basados en el Teorema de Bayes, para la predicción de calificaciones o notas ante diferentes tipos de incertidumbre: a) *a priori* objetiva, cuando se dispone de datos expresados en registros de calificaciones anteriores de puntuaciones de una prueba. b) *A priori* subjetiva, si se dispone de datos cualitativos sobre el desempeño del estudiante o si el profesor expresa sus creencias sobre el futuro desempeño del estudiante. c) *A priori* no informativa, en el caso de que no se tenga información alguna sobre el desempeño anterior del alumno.

¿Cuáles son las consideraciones que deben realizarse para predecir la calificación final de un estudiante durante un curso? ¿Los datos proporcionados por las puntuaciones de sus pruebas?, o ¿las expectativas del profesor sobre su futuro desempeño? ¿Acaso sería deseable estimar la calificación comparando el puntaje obtenido, por

este alumno en particular, en una prueba con normas o estándares predeterminados cuando este mismo instrumento de medición se aplicó a una muestra de estudiantes? Este estudio pretende dar respuesta a estas interrogantes y proporcionar un instrumento útil para el profesor y el administrador educativo para tomar acciones correctivas antes de que termine el ciclo escolar y de que el único camino que subsista sea el que lleve a reprobar a un gran número de estudiantes.

Ahora bien, el profesor debe actuar ante información diversa e insuficiente y si espera a tener los suficientes datos objetivos, tal vez la reprobación, sea irreversible. De ahí la posible utilidad de los procedimientos que se proponen.

¿Cuál será la forma en que se demostrará la utilidad de los procedimientos bayesianos propuestos? Siendo uno de los objetivos más importantes de la ciencia la predicción de los fenómenos de los que se ocupa, la utilidad se demostrará, precisamente, al producir inferencias válidas a partir de los datos disponibles, ya sean estos objetivos o subjetivos. La forma de lograrlo será la siguiente: las inferencias resultantes de los tres procedimientos serán comparadas (con un procedimiento de prueba de hipótesis apropiada al tipo de mediciones realizadas) con la realidad: los datos de la población, lo que permitirá

apreciar cuál de los tres procedimientos es el más certero en cuanto a grado de exactitud en la predicción se refiere.

Con el fin de enmarcar la evaluación referida a un criterio, se mencionarán algunos hechos considerados relevantes en la historia de la medición y evaluación psicológicas, particularmente las primeras pruebas y calificaciones, enseguida, se plantearán los antecedentes propiamente dichos, que iniciarán con una revisión de los conceptos de medición, niveles de medición y evaluación desde el punto de vista de las pruebas de aprovechamiento escolar y de la asignación de calificaciones o notas.

A continuación se analizará el problema de la asignación de calificaciones o notas previo a un análisis de las ventajas y desventajas que tendría dicha asignación si es realizada con referencia a una norma o a un criterio especificado en los objetivos de aprendizaje de la materia.

Este trabajo se encuentra enmarcado precisamente dentro de los principios de la evaluación referida a un criterio y pretende ser utilizado, al predecir las notas, esto es el futuro desempeño del alumno,

como un instrumento de evaluación formativa y toma de decisiones ante diferentes tipos de información disponibles.

El curso de la argumentación seguirá con la reseña de estudios recientes de aplicación de la estadística bayesiana en procedimientos de evaluación formativa y toma de decisiones ante diferentes tipos de incertidumbre, así como con argumentos a favor y en contra de la utilización de criterios subjetivos y objetivos en evaluación educativa y particularmente en asignación de calificaciones.

Finalmente, se procederá al planteamiento del problema de estimación de calificaciones ante diversos tipos de información disponible relacionándolo con las posiciones divergentes al respecto y postulando los procedimientos propuestos como una posible respuesta al dilema de la objetividad *versus* la subjetividad.

### ***Antecedentes de la medición y evaluación psicológicas***

El origen de las pruebas psicológicas se remonta probablemente al siglo XII a. C. cuando el emperador chino reinante de la dinastía Chan ordenó que se sometiera a prueba a sus oficiales de gobierno para

que se estableciera de este modo su capacidad para el cargo. (Gregory, 2001), (Cohen, 2000).

Por su parte, Díaz Barriga, (2000), afirma que la prueba no es un elemento inherente a la educación y que fue creada por la burocracia china para elegir miembros de castas inferiores.

Los exámenes por escrito se introdujeron en la dinastía Han (202 a.C. – 200 d.C.) como resultado, probablemente, de la evolución y perfeccionamiento de las pruebas. A partir de 1370 y hasta 1906, en que se abolieron por orden real debido al descontento, las personas que deseaban obtener el puesto público de mandarín, debían someterse a una serie de pruebas con criterios específicos muy rigurosos, tales como redactar ensayos, escribir poemas, saber manejar el arco y la flecha y equitación. Cohen, (2000) afirma que en ocasiones los examinados morían debido a la tensión producida por las pruebas aplicadas. Las condiciones en que se desarrollaban los exámenes requerían hasta de tres días con sus noches, y de cada sesión de pruebas aprobaban entre el 1 y el 3% de los aplicantes.

¿Cuáles eran las razones por las cuales fue diseñado este sistema de exámenes? Para Judges, (2000), la tradición China de varios

milenios de exámenes literarios basados en los escritos canónicos de Confucio para la selección de funcionarios sobre la base de su habilidad demostrada (prueba referida a un criterio), dejaba de lado una élite hereditaria y actuaba como un estabilizador de los miembros de la jerarquía del poder. Piensa el autor que los rasgos de este sistema de exámenes contienen algunos elementos coincidentes con el desarrollo de la práctica de exámenes en otras partes: el poder de un sistema con los propósitos de mantener la cultura y la tradición, funcionar como medio de control social, poder para racionalizar las funciones humanas y las normas de competencia, discriminar la inteligencia y dividir a una población con base en diferentes *status*.

En la Biblia, se consigna la aplicación de una evaluación referida a un criterio, cuyo incumplimiento traía como consecuencia la muerte: Jefé, Galaita, vence a los Efraím que se habían sublevado y colocándose su gente en los lados del Jordán por donde pasarían los de Efraím, al pasar un transeúnte le pedían que pronunciara la palabra "schibbolet" (que significa espiga) y si el caminante pronunciaba "sibolet", lo degollaban.

De acuerdo con el análisis desarrollado por Judges, (2000), en la antigua Grecia y en Roma, los exámenes no representaban un papel de

selección. A los jóvenes de la clase gobernante se les ejercitaba en el arte de la dirección pública y de la elocuencia persuasiva en el lenguaje hablado y escrito, pero, a pesar de que en la Roma Imperial la habilidad en la retórica influía para la admisión en puestos de gobierno, se desconocen listas de aprobados o sistemas de calificaciones.

Un ejemplo de evaluación referida a un criterio muy estricto en el dominio psicomotor, donde se considera que cumple ampliamente sus funciones, Stanley, y Hopkins, (1972) se da en la Edad Media en los gremios o corporaciones, en donde el aprendiz debía realizar una obra de la misma calidad que la del maestro del taller, y este hecho se consideraba indispensable para que el estudiante fuera considerado ya como maestro.

De acuerdo con Judges, (2000), en Europa, en el siglo XII, en Bolonia, París y posteriormente en Oxford, resurgió el estudio del *trivium*<sup>1</sup> y el *quadrivium*<sup>2</sup> (las siete ciencias exactas) y se incluyó un proceso de investigación sobre la idoneidad de los candidatos; sin embargo, para el otorgamiento de títulos y grados, la prueba de competencia se realizaba frecuentemente de manera informal y sin el requisito de un trabajo escrito.

---

<sup>1</sup> *Trivium*. Gramática, retórica y lógica.

<sup>2</sup> *Quadrivium*. Gramática, retórica, lógica, geometría, aritmética, música y astronomía.

Para Durkheim, (1938), los grados y los exámenes no aparecen en el mundo antiguo sino hasta la aparición de las universidades en la Edad Media, de hecho, no existe una palabra en latín que signifique claramente examen. El sistema jerárquico de grados aparece con sus exámenes ritos y ceremonias por una influencia del sistema de los gremios o corporaciones que siendo una especie de sociedad secreta del Medievo, en donde los aspirantes pasaban una serie de pruebas y juramentos, somete a los estudiantes, a través de los grados, a una “iniciación” paulatina. Sin embargo, prácticamente no había “reprobados” ya que existen registros de que el número de admitidos y el de candidatos era el mismo, pero aproximadamente del número de estudiantes inscritos sólo la mitad terminaba el bachillerato y menos de la mitad alcanzaba la licenciatura. Esto lleva a concluir que únicamente se admitía al examen de grado a aquellos que se consideraba eran aptos para “aprobarlo”.

Juan Amos Comenio en su obra “Didáctica Magna” en 1657, considera que el examen no es un elemento que defina la promoción ni una calificación expresada en número. Asimismo, proporciona numerosas sugerencias para que el maestro realice evaluaciones formativas, informales en el sentido actual, y sugiere que se pregunte a

los alumnos: ¿Con qué motivo hemos llegado a esto?, ¿qué acabo de decir?, y si el alumno interrogado dudaba, Comenio sugería pasar a otro y si fuese necesario al trigésimo. Sugiere preguntar las lecciones, que en esos tiempos se aprendían de memoria, preguntando un párrafo a un estudiante y el siguiente al otro y así sucesivamente. Los ejercicios de traducción se corregían a todos con ayuda de los Decuriones, una especie de maestros auxiliares que corregían la ortografía. Los alumnos corregían así los errores que encontraban en las traducciones de sus compañeros. Así, Comenio proporciona una gran cantidad de formas en que un profesor con cien alumnos podría realizar evaluaciones formativas que coadyuvaran con el “aprendizaje”. Actualmente, se podrían establecer, con base en los objetivos de aprendizaje, listas de conductas a considerar y se podría utilizar la estadística bayesiana para la realización de las predicciones sobre el futuro desempeño del alumno.

En Alemania, en tiempos de Federico el Grande<sup>3</sup>, se instituyó el *abitur*, examen final de la escuela clásica y en el siglo XX otro, el *abiturientenexamen* que tuvo como consecuencia el otorgamiento del título de *bachelor* de las universidades.

---

<sup>3</sup> Federico II, El Grande, de Prusia, 1831-1888.

Gregory, (2001) realiza una explicación pormenorizada de la primera prueba psicológica desarrollada y adaptada por Binet y Simón en la Sorbona en 1905, 1908 y 1911. El objetivo de este instrumento fue detectar a los niños que requerían educación especial en las escuelas. Se pretendía separarlos y proporcionarles el tipo de educación que necesitaban, lo cual puede ser considerado tangencialmente como evaluación formativa referida a un criterio. Los criterios fueron, para el instrumento de 1905, diferentes habilidades y conocimientos que iban desde seguir un objeto en movimiento hasta definir palabras abstractas. La revisión de 1911 contenía para cada nivel de edad 5 pruebas. Binet y Simón estandarizaron sus pruebas e introdujeron el concepto de nivel mental que pronto se transformó en edad mental.

### ***Medición y niveles de medición***

Medir, del latín *metiri* término definido en el diccionario como la determinación de la longitud, extensión, o capacidad de algo y como la determinación de la cantidad de una magnitud por comparación con otra que se toma como unidad (1999), también se concibe en el Diccionario Enciclopédico Olimpia (2000) como la comparación de una cantidad con su respectiva unidad con el fin de averiguar cuántas veces la segunda está en la primera.

Las tres definiciones anteriores involucran que el tamaño de los intervalos en las escalas de medición sean iguales, lo que no sucede con frecuencia en las pruebas de rendimiento escolar, que son un factor importante y en algunos casos el único, en la asignación de calificaciones.

Por su parte, Glass, y Stanley, (1970) consideran que la medición es la asignación de números a ciertos objetos de acuerdo con determinadas reglas y, si es ésta la forma en que transformamos en números ciertos atributos de nuestras percepciones, esta definición de medición podría concordar con la asignación de probabilidades objetivas y subjetivas.

Ahora bien, la medición tiene diferentes niveles que se encuentran representados por diferentes escalas. Glass, y Stanley, (1970) consignan que existen casos en los que la medición se limita a determinar si existe o no el atributo y es imposible que una sola observación se encuentre en dos categorías al mismo tiempo, este es el caso del nivel de medición nominal. Un ejemplo sería asignar el número 1 al hombre y el 2 a la mujer. También podría considerarse en esta categoría la calificación de aprobado o reprobado que se otorga en

las clases de laboratorio del Instituto Politécnico Nacional, aunque se podría argumentar también, que estarían en un nivel ordinal con dos niveles.

Al siguiente nivel de medición le corresponde la escala ordinal, en la cual se considera que se poseen diferentes cantidades del mismo atributo, pero se desconoce con precisión qué tanto más o menos se posee de éste, ya que el tamaño de los intervalos no es igual. Éste es el caso de la mayoría de las pruebas de aprovechamiento que se aplican en las escuelas. Se diseña y aplica una prueba dándole valores arbitrarios a cada pregunta, pero no se tiene la certidumbre de que los dos puntos que “vale” una respuesta son equivalentes a otros dos puntos otorgados a otra respuesta.

La escala de intervalos corresponde al siguiente nivel de medición. En este caso, es posible determinar la cantidad del atributo en una sola observación ya que el tamaño de los intervalos es igual. Es el caso de las escalas centígrada y Fahrenheit para medir la temperatura, en la primera se fijó arbitrariamente el punto cero a la temperatura de fusión del agua y  $100^{\circ}$  C a la que corresponde con el punto de ebullición del agua al nivel del mar.

El desarrollo de la tecnología de construcción y análisis estadístico de los *tests* hace que cuando se asignan calificaciones utilizando preguntas de opción múltiple y como referencia la curva normal, se considere que se está trabajando con una escala de intervalos, Magnusson, (1966), sin embargo, tal vez se esté trabajando en una nueva escala, entre ordinal e intervalar, ya que las calificaciones que emergen de dos pruebas de aprovechamiento, aunque se pueden comparar, estadísticamente hablando, no es posible considerar que tienen el mismo significado que la transformación de grados centígrados a Fahrenheit y viceversa.

La última escala de medición es la de razón, y en ésta, además de que el tamaño de los intervalos es igual, se tiene un punto cero que indica la carencia total del atributo a medir, así, la escala Kelvin de temperatura establece como  $0^{\circ}$  K el cero absoluto de temperatura, esto es, la ausencia total de movimiento molecular. Los tiempos de reacción serían una observación, utilizada frecuentemente en psicología, para este nivel de medición.

La asignación de calificaciones, en la mayoría de los casos, puede ser considerada como un proceso de medición exclusivamente, ya que muchos profesores se limitan a asignar puntos a las pruebas y,

posteriormente, la suma de ellos equivale a la calificación asignada. En este caso, el proceso de medición se equipara al de asignación de calificaciones y el tipo de escala de medición utilizada sería la ordinal.

### ***Evaluación y sus funciones***

El término evaluar es considerado en el diccionario Larousse, (1999) como la acción de tasar, valorar, atribuir el valor a algo, también se encuentra una definición que involucra las labores de los profesores y así consigna que evaluar es estimar los conocimientos y rendimiento de un alumno. Cuando se está evaluando, se está también, de hecho, realizando una predicción sobre el futuro desempeño del estudiante. Así, la práctica de apreciar, estimar, tasar, valorar, justipreciar, juzgar y atribuir el valor a algo, que puede o no realizarse con base en los resultados de un examen y/o de las apreciaciones subjetivas de los maestros, sería equiparable a la de asignación de calificaciones. Evaluar podría confundirse en determinadas circunstancias con medir, pero, lo que diferencia a este proceso del de la medición, es la entrada de la subjetividad en gran medida, ya que para justipreciar es necesario algo más que la simple asignación de números.

Completando las definiciones anteriores, Gimeno (1999) considera que la evaluación educativa es, además, una práctica que se desarrolla siguiendo determinados usos, que cumple diversas funciones y cuya realización se ve condicionada por un sinnúmero de aspectos personales, sociales e institucionales y cuyas repercusiones tocan el orden de lo personal, lo institucional, lo familiar y lo cultural. La evaluación puede verse desde dos puntos de vista, su carácter didáctico o desde el punto de vista crítico por la multitud de fenómenos que desencadena. Para Gimeno: "Evaluar hace referencia a cualquier proceso por medio del que alguna o varias características de un alumno, de un grupo de estudiantes, de un ambiente educativo, de objetivos educativos, de materiales, profesores, programas, etc., reciben la atención del que evalúa, se analizan y se valoran sus características y condiciones en función de unos criterios o puntos de referencia para emitir un juicio que sea relevante para la educación". (p. 3). Esta definición involucra la evaluación con referencia a un criterio y la evaluación con referencia a la norma. El presente trabajo se encuentra encuadrado en la evaluación referida a un criterio.

Desde el punto de vista didáctico, Bloom, Thomas, y Manaus, (1975) definen los términos de evaluación diagnóstica, formativa y sumativa. Se entiende por evaluación diagnóstica al proceso de medir e

interpretar las mediciones antes de iniciar un ciclo de enseñanza. Una acción posible sería modificar el programa o plantear alguna acción remedial, en el caso de que los estudiantes desconozcan los requisitos para cursar con éxito la asignatura. Otra acción se tomaría si los estudiantes conocieran parte o la totalidad de la materia, para lo cual se haría necesario modificar el programa de estudio o la profundidad con que se toquen los temas del mismo. ¿Es posible aplicar el procedimiento que se propone en este trabajo en el caso de evaluaciones diagnósticas? Sí, y además de considerar los resultados de la prueba, se podrían considerar otros datos disponibles como promedio de calificaciones en todas o en algún grupo de asignaturas, tendencias de calificaciones en esa asignatura en periodos previos y valoraciones subjetivas en el caso de que el profesor ya conozca a los alumnos.

Evaluación formativa será aquella que se realiza en algún punto del proceso y que tiene por objetivo el realizar acciones correctivas que lleven a que el alumno culmine con éxito el ciclo escolar. Éstas podrían ser de diversa índole, en función del nivel alcanzado y de las deficiencias detectadas.

Evaluación sumativa es la que se lleva a cabo al finalizar el período de instrucción y tiene como fin determinar qué tan bien se

cumplieron los objetivos de aprendizaje. Los procedimientos de estimación de calificaciones propuestos pueden ser utilizados en los tres casos anteriores ya que esto depende del momento en que se realice la evaluación, del propósito de la misma y de las decisiones que se tomen en consecuencia.

En el caso de la evaluación referida a una norma, a la puntuación en bruto obtenida por un estudiante, se le realizan diversos tratamientos estadísticos y se compara con las normas de puntuaciones encontradas en una muestra muy amplia de individuos que resolvieron la prueba, lo que permite valorar la puntuación en particular. En el caso de evaluación referida a un criterio, éste se encuentra especificado en los objetivos de la materia, redactados en términos conductuales, con un criterio de ejecución aceptable y describiendo las condiciones en que se deberá llevar a cabo el examen. Mager, (1965).

### ***Pruebas referidas a una norma***

Resulta necesario definir en primer lugar qué se entiende por prueba. Gregory, (2001), dice: "Una prueba es un procedimiento estandarizado para tomar una muestra de conducta y describirla con categorías o puntuaciones. Además, la mayor parte tiene normas o

estándares, con base en los cuales pueden utilizarse los resultados para pronosticar otras conductas más importantes” (p. 147). Esta definición corresponde básicamente a una prueba referida a la norma.

La evaluación referida a una norma se realiza partiendo del diseño y construcción de una prueba confiable y válida que se perfecciona a través de pilotajes y que en su versión definitiva se aplica a una muestra amplia de individuos que cumplen las características de la población para la que fue diseñada la prueba. Las puntuaciones de la prueba reciben un tratamiento estadístico de normalización que convierte los puntajes en bruto en calificaciones estándar que hacen posible la comparación de los puntajes individuales en función de que se cuenta con datos precisos de qué porcentaje de la población obtuvo puntajes similares y de qué lugar relativo (a la derecha o a la izquierda de la media) ocupa una medición particular.

Existen diferentes propuestas con relación a la utilización de este procedimiento para asignar calificaciones o notas. Otorgar la calificación más alta a los puntajes de aquellos alumnos que ocupen el rango más elevado a la derecha de la curva y reprobar a los que se encuentren a la izquierda de la media o aproximadamente al siete por ciento de la población. A pesar de los intentos de eliminar la

subjetividad a través de la construcción de pruebas objetivas con puntajes normalizados, los criterios para asignar notas con referencia a las norma dependen del juicio subjetivo del profesor.

Según Gimeno (1999), bajo la óptica del positivismo, se ha sobrevalorado la importancia de los métodos para constatar objetivamente el rendimiento del alumno, por encima de su capacidad para proporcionar información a los profesores ya que en la búsqueda de la objetividad, al normalizar las puntuaciones para que sirvan para comparar los resultados de diversos sujetos, se ha empobrecido la enseñanza, ya que las preguntas de las pruebas objetivas raramente elicitán procesos intelectuales complejos y anulan la expresión personal de los estudiantes. ¿Cuáles son las razones por las que los estudiantes memorizan los temas, en lugar de buscar niveles más elevados que los del simple reconocimiento y discriminación? Tal vez haya contribuido a lo anterior la falta de creatividad de muchos elaboradores de pruebas de opción múltiple.

Por otro lado, con la consigna de que deben reprobar los alumnos que se encuentren en el extremo de puntuaciones alcanzadas más bajas, se puede tener el caso de un grupo de alumnos excepcionales que han alcanzado el dominio de los objetivos de la materia en un

porcentaje, por ejemplo, del 85% como mínimo y que tengan que ser reprobados.

### ***Pruebas referidas a un criterio***

Una prueba referida a un criterio es una forma de proceder predeterminada que a partir de la toma de una muestra de conducta, atingente a un tema y a los objetivos de aprendizaje correspondientes, predice la conducta ulterior de la persona con base en el dominio demostrado de los objetivos.

La asignación de calificaciones basada en la evaluación referida a un criterio está sustentada en la consecución de las conductas descritas en los objetivos de la materia, bajo los criterios de ejecución aceptables y las condiciones especificadas en los mismos. Así, una forma de asignar calificaciones podría ser la siguiente: si alguna persona contesta su examen totalmente, de acuerdo con los objetivos de la materia, obtiene la calificación más elevada y aquella persona que conteste el 70 % de la prueba obtendrá la calificación equivalente a este porcentaje.

¿Qué procedimiento será conveniente utilizar cuando se trata de pruebas referidas a un criterio, esto es, el criterio especificado en los objetivos de aprendizaje? ¿Qué procedimiento será conveniente utilizar cuando los reactivos de examen son los correspondientes a solución de problemas?

Los procedimientos que se proponen en este estudio, no interpretan o juzgan el desempeño en función de la posición relativa con referencia a la “norma de grupo” sino que es interpretado con respecto a un criterio previamente establecido. El alumno debe demostrar su pericia en resolver problemas de aplicación de fisicoquímica, con un procedimiento apropiado y con una exactitud numérica con una desviación real no mayor al 10%. Se está tratando el caso de evaluación referida a un criterio. Sin embargo, Cohen, (2000), afirma que aún en el caso de que el resultado de la evaluación sea de carácter nominal: aprobado-reprobado, existe un reconocimiento implícito de que existe un continuo de capacidad y que se ha decidido establecer un punto de corte, por lo que en sentido estricto, las pruebas referidas a un criterio son también normativas.

Cohen (2000) considera que las pruebas referidas a un criterio se utilizan ampliamente en programas educativos asistidos por

computadora, donde se realiza evaluación formativa del desempeño en un segmento de material y se toma la decisión de si el alumno puede seguir con el siguiente nivel de instrucción o repetir el segmento anterior con los mismos o diferentes materiales.

La evaluación referida a un criterio también se aplica para investigar si los estudiantes dominan diferentes actividades académicas, por ejemplo la lectura, las matemáticas, etc.; en los exámenes profesionales en donde se deciden ciertos criterios de desempeño para obtener el título de una profesión; en las escuelas, donde se tiene que cumplir con un plan de estudios y para evaluar el dominio de destrezas tales como soldar, manejar un avión, saber operar un equipo, por ejemplo.

***Asignación de calificaciones referida a un criterio Vs.  
evaluación referida a la norma***

¿Cuál es la mejor forma de asignar calificaciones? ¿La referida a un criterio o la referida a la norma? Para Díaz Barriga ( 2000), la lógica de la construcción de las pruebas conduce a arbitrarios

epistemológicos, así, los pasos: 1) atributo, 2) operacionalización del atributo y 3) construcción del instrumento, carecen de un paso intermedio entre el primero y el segundo que es una construcción conceptual que resulte comprensiva del fenómeno dejando al juicio de quien diseña el *test* la determinación de las operaciones que a su juicio reconoce como válidas para medir el atributo: inteligencia o aprendizaje.

Sin embargo, en el caso que nos ocupa, que está inserto en la evaluación referida a un criterio, los objetivos de la materia de Fisicoquímica III y de la mayoría de las materias de la carrera de Ingeniería Química Industrial en la ESIQIE, IPN, consisten en resolver problemas de aplicación y los exámenes se diseñan precisamente con problemas de aplicación. El arbitrario epistemológico señalado por Díaz Barriga no es aplicable para este caso.

### ***Estudios bayesianos en evaluación educativa:***

#### ***Objetividad vs. subjetividad***

Con referencia a las líneas de interpretación de la probabilidad, Thomas Bayes en su obra: "Essay Towards Solving a Problem in the Doctrine of Chances" (1763) además de establecer una fundamentación

de la inferencia estadística, proporcionó los principios lógicos para inferir: dada la probabilidad de un evento, las probabilidades de eventos relacionados. Para Bayes, la probabilidad de cualquier evento es: "...la razón entre el valor al cual una *expectativa* que depende de la ocurrencia de un evento, debe ser computada y el valor de la cosa esperada sobre su ocurrencia" (p. X).

Basado en el enunciado de la Propiedad 2 del documento de Bayes (p. 299) en donde se enuncia que la expectativa de una persona sobre la ocurrencia de un evento, esto es su *creencia personal*, es una probabilidad previa que se toma en cuenta para el cálculo de probabilidades condicionales, Savage (1954), rechaza la probabilidad objetiva e interpreta a la probabilidad como el grado de creencia personal de un individuo sobre la ocurrencia de un evento; de esta manera, la probabilidad *a priori* de una persona, es su creencia antes de que se realicen observaciones y la probabilidad *a posteriori* es aquella que tendrá la persona después de que las observaciones hayan sido hechas. Por otra parte, los estudios de probabilidad están basados en datos objetivos obtenidos de muestras, así, Jeffrey (1961), afirmó que aún para la total ignorancia existe una distribución objetiva de grados de confianza para varias hipótesis y desarrolló una regla para estimar probabilidades *a priori* no informativas. Savage y Jeffreys

representan posiciones divergentes entre los estadísticos bayesianos contemporáneos.

¿Cuándo se está hablando de un argumento objetivo y en qué ocasiones de uno subjetivo? El problema de la objetividad y subjetividad es tan antiguo como la filosofía y tan moderno como la física cuántica. Platón<sup>4</sup>(s. IV a. C. /1952) en el mito de la caverna, en su República, representa simbólicamente la relación del hombre y su subjetividad con la filosofía y con la realidad. Los hombres que se encuentran desde niños en una caverna que tiene una abertura por donde entra la luz exterior, pero están sujetos de manera que sólo pueden mirar al fondo donde la luz de un fuego encendido hace que se reflejen sombras provocadas por todo tipo de objetos y estatuillas que portan otros hombres. La caverna es el mundo sensible y objetivo y el mundo “real” es el de la razón y las ideas. Marías (1941) al analizar la teoría de Platón sobre la realidad la divide en dos partes, el mundo sensible (realidad aparente) y el mundo inteligible (realidad verdadera), en el primero se encuentran las sombras, las conjeturas, las cosas reales y las creencias y en el segundo los objetos matemáticos, las ideas y el discurso. Para Platón la realidad verdadera es el pensamiento y el mundo objetivo es un mundo de sombras. Una *a priori* subjetiva, fruto

---

<sup>4</sup> 428 -348 a. C.

del razonamiento en el mundo de la “realidad verdadera” de las ideas sería tal vez para Platón la más acertada, ahora bien el mundo sensible que es conocido como la realidad objetiva, para Platón, es tan solo un simple reflejo o sombra de las cosas reales, probablemente por la limitación de los órganos de los sentidos, así, la *a priori* objetiva, fruto de nuestras observaciones, sería sólo una sombra de la realidad.

Descartes<sup>5</sup>, en la primera mitad del siglo XVII, considera que nada le merece su confianza ya que en el pasado las opiniones de los filósofos se contradicen. Los sentidos nos engañan con frecuencia, además, ¿cuál es la realidad del sueño y la alucinación?; tampoco se puede confiar en el pensamiento ya que frecuentemente cae en el error, únicamente la lógica y las matemáticas son ciencias reales que nos ayudan a conocer la realidad. Pero, qué se puede hacer ante tal incertidumbre, dice Descartes que hay que dudar de todas las cosas. Funda su especulación en el criterio de evidencia, pero esta evidencia no la proporciona el mundo objetivo, pues es percibido por los sentidos que nos engañan con frecuencia, sino el mundo de la razón. Así, dice “En los sujetos que nos proponemos investigar, nuestras indagatorias deben ser directas, no lo que otros hayan pensado ni lo que nosotros hayamos conjeturado, sino lo que nosotros podemos clara y

---

<sup>5</sup> René Descartes, nació en 1596 y murió en 1650.

perspicazmente sostener y con certeza deducir; el conocimiento no se puede obtener de ninguna otra manera". (p. 3). Así para Descartes, la *a priori* subjetiva que está basada en una serie de razonamientos, sería la adecuada y la objetiva, basada en observaciones realizadas por nuestros órganos de los sentidos que nos engañan, no sería un criterio utilizable para construir el conocimiento.

Según Marías, (1944), el racionalismo cartesiano es antecedente del apriorismo y posiblemente influyó en el pensamiento de Thomas Bayes y su concepto de *a priori* subjetiva, ya que se considera que la única instancia de valor para el hombre es la razón.

Francis Bacon<sup>6</sup> fue el iniciador de una nueva corriente filosófica: el empirismo e introdujo el método inductivo, que es una postura diametralmente opuesta al idealismo racionalista y matemático representado por Descartes. En el *Novum Organon*, afirma que es el ser humano el ministro e intérprete de la naturaleza, quien entiende por medio de sus observaciones el orden de la naturaleza y son el hombre y las observaciones aunados a las "cosas" de la mente lo que le permite conocer. Para Bacon, de una colección de hechos individuales, agrupados de modo sistemático y conveniente, se obtienen por

---

<sup>6</sup> Francis Bacon, nació en 1561 y murió en 1621.

abstracción, después de seguir un proceso experimental lógico riguroso, los conceptos generales de las cosas y las leyes de la naturaleza.<sup>7</sup> Dice Bacon que la sutilidad de la naturaleza está más allá del sentido o del entendimiento así es que las meditaciones, especulaciones y teorías sobre la humanidad no son otra cosa que una locura ya que no existe quien esté allí y lo observe. También afirma que los términos sustancia, cualidad, acción, pasión y existencia no son nociones claras y que es indispensable definir otros como peso, densidad, material, elemento, etc. Definitivamente, Bacon rechazaría totalmente una *a priori* subjetiva por provenir de la intuición y tal vez del razonamiento o las matemáticas (que Bacon consideraba dentro del conjunto de los magos y alquímicos con éxitos “magros”), sin concurso de la observación. Así, desde el punto de vista empírico la *a priori* objetiva sería la basada en las observaciones sistemáticas de la naturaleza y la *a priori* subjetiva, la sustentada en consideraciones desligadas de la observación y por lo tanto sinónimo de locura.

Hume (1748/1852) sigue la tradición empirista. Para este filósofo las ideas que atribuimos a la sustancia se fundan en una impresión intuitiva. La percepción y la reflexión nos proporcionan una serie de elementos que atribuimos a la sustancia. Así, yo (siguiendo el estilo

---

<sup>7</sup> Bacon, F. (1952, Reimpreso ) *Novum Organon, Aphorisms Concerning the Interpretation of Nature and the Kingdom of Man.* , en *The Great Books, Volumen 30, Encyclopaedia Britannica, Inc.: The University of Chicago, Estados Unidos de América* ,107-195.

explicativo de Hume) encuentro las impresiones de color, dureza, sabor, olor y extensión, figura redonda y suavidad y lo refiero a un todo desconocido que llamo manzana, esa es una sustancia y como las impresiones sensibles son más vivas que las imaginadas yo tengo la creencia ( *belief*) que eso es la realidad pero no puedo estar seguro de ella, ni siquiera de la existencia del propio yo que también es una colección de percepciones o contenidos de conciencia. De esta forma, el empirismo de Hume se convierte en escepticismo y afirma que la realidad es sólo una percepción, es experiencia, es idea, todo son impresiones subjetivas. Con referencia a la probabilidad, para Hume no existe algo llamado suerte en el mundo y es nuestra ignorancia sobre la causa real de cualquier evento la que tiene una influencia sobre el entendimiento y se expresa como una creencia u opinión. Así, para David Hume, toda *a priori* y toda *a posteriori* serán, necesariamente, subjetivas.

Dice Kant en la Crítica de la Razón pura (Kant, 1883/1979) que si bien todo conocimiento comienza con la experiencia, no todo el conocimiento se origina en la experiencia ya que este conocimiento tiene dos componentes: lo que recibimos de fuera por medio de impresiones sensibles y lo que nuestra propia facultad de conocer

aporta. Para Kant<sup>8</sup>, conocer es una función activa del sujeto, no es recibir algo que está allí, sino hacer algo que se conoce, poner algo. De esta forma Kant dice que conocemos de las cosas lo que nosotros hemos puesto, por lo que las cosas no están allí, sino que las hago yo al conocerlas (Marías, J., 1941) Para Hume, es imposible todo conocimiento objetivo, Kant afirma que si bien no podemos saber cómo son las cosas en sí mismas, si podemos saber de ellas en cuanto a como se nos presentan: en cuanto fenómenos; nosotros ponemos algo, los conceptos del espacio y del tiempo no están en las cosas sino en nuestra razón. De la misma, forma Einstein (1960/1986) considera que el tiempo y el espacio son esquemas con arreglo a los cuales pensamos y no condiciones en las que vivimos. Dice: "Lo primero que hacemos es suponer que existe un orden temporal de los acontecimientos, que concuerda con el orden temporal de las experiencias. Esto es lo que generalmente hacíamos de un modo inconsciente, hasta que se suscitaron ciertas dudas escépticas" (p. 185) Comenta que el orden temporal de las experiencias obtenido, por ejemplo, por la vía acústica, puede diferir del orden temporal obtenido por la vía visual, de tal forma que el orden temporal de los acontecimientos no se puede identificar simplemente con el orden temporal de las experiencias personales. Einstein considera que los conceptos de espacio, tiempo y

---

<sup>8</sup> 1724-1804

acontecimiento, que pretendemos llamar objetivos, de “naturaleza espacial”, en oposición con los conceptos subjetivos de la esfera psicológica tienen su origen en los conceptos de objeto corpóreo y cuerpo rígido que son de naturaleza precientífica y subjetiva. Einstein considera que es necesario hacer descender los conceptos fundamentales del pensamiento científico, “desde las esferas olímpicas de Platón” para tratar de descubrir su origen terrestre y liberar así a esos conceptos del tabú que se les asocia. Considera también que David Hume tiene el mérito de haber realizado esta reflexión crítica.

Para Kant, los conocimientos aparentemente desligados de la experiencia son los *a priori* y se distinguen de los empíricos en que estos tienen sus fuentes *a posteriori*, en la experiencia. Sin embargo, siguiendo a Kant, esos conocimientos que aparentemente están desligados de la experiencia inmediata, fueron también sacados de la experiencia. Para Kant los problemas de la razón pura totalmente *a priori* serían Dios, la libertad y la inmortalidad; y las matemáticas serían un excelente ejemplo de cuán lejos es posible ir en el conocimiento *a priori*, independientemente de la experiencia. La labor de la razón consiste en el análisis de los conceptos que ya tenemos de los objetos. Así para Kant, una *a priori* subjetiva, basada en creencias o razonamiento tendría un sustrato objetivo, ya que esas creencias

estarían sustentadas en experiencias previas y una *a priori* objetiva también tendría un componente subjetivo, pues la percepción realizada a través de los sentidos integraría una serie de constructos con arreglo a los cuales pensamos (siguiendo también a Einstein) y no condiciones “reales” en las que vivimos.

Marías (1941), afirma que para Hegel, los filósofos que dicen poseer la verdad porque les fue revelada por Dios han convertido a la filosofía en nebulosidad. Opina que la verdadera figura en la cual existe la verdad no puede ser sino el sistema científico de la misma. Colaborar a que la filosofía se aproxime a la forma de ciencia –a que pueda despojarse de su nombre de amor al saber y sea saber efectivo– es lo se había propuesto. La realidad es el absoluto que existe en una evolución dialéctica de carácter lógico racional, Hegel considera que todo lo real es racional y que todo lo racional es real. Hegel reconoce un conocimiento conceptual, en el cual se tienen conceptos de las cosas y ese conocimiento serían las ciencias, y en la búsqueda del saber absoluto que no puede dejar nada fuera de sí, ni siquiera al error, desarrolla una lógica y ontología de la razón que se sustenta en el concepto de espíritu que es el ser para mí, mismidad, la entrada en sí mismo, la mismidad, el ser para sí. Sin embargo, Hegel define al espíritu subjetivo en cuanto está unido a un cuerpo, convirtiéndose en

este momento en alma y es objeto de estudio de la antropología, y en cuanto espíritu, objeto de estudio de la psicología y al estudio del espíritu objetivo, que estaría allí sin estar unido a un cuerpo, pero que sería subjetivo desde el punto de vista de la definición de espíritu (mismidad), se dedicaría el derecho, la moral y la eticidad. Así, desde el punto de vista de Hegel la psicología en sí misma está dedicada al estudio del espíritu subjetivo y, tal vez, si una *a priori* estuviese sustentada racionalmente, aún sin concurso de datos producto de observaciones, y dado que toda realidad es racional y lo racional es real, sería igualmente válida que una *a priori* sustentada en mediciones.

Después del idealismo de Hegel, surge la necesidad de apartarse de las explicaciones mentales y atenerse a las cosas, a la realidad misma, lo cual resulta inobjetable, pero, ¿cuál es la realidad? Para el positivismo, la realidad son los hechos sensibles, pero esto es un error, pues la realidad es diferente a lo que aparece a nuestros sentidos, a los fenómenos y no es posible creer que algo “real” es la “realidad”.

Según Marías (1941), Comte<sup>9</sup> considera que los conocimientos pasan por tres estados teóricos distintos y desarrolla la ley de los tres estados, base de la filosofía y psicología positivista. En el primero,

---

<sup>9</sup> 1798-1857

teológico o ficticio, la mente busca las causas y los principios de las cosas y tiene tres estadios: el fetichismo, el politeísmo y el monoteísmo. El segundo es el metafísico o abstracto en el que se siguen buscando los conocimientos absolutos a través de la construcción teórica de entidades abstractas que expliquen. Así la Naturaleza es la gran entidad general que sustituye al concepto de Dios. Y el tercer estado es el positivo o real y es el definitivo ya que en él la imaginación es sustituida por la observación. El positivismo busca únicamente hechos y sus leyes, ni causas, ni principios que son inaccesibles. El positivismo es relativo ya que el estudio de los fenómenos es relativo a nuestra organización y a nuestra situación. Del positivismo surge el conductismo para el que la psicología es una rama de las ciencias naturales y la objetividad, la predicción y el control son requisitos esenciales en la producción del conocimiento científico. A partir del modelo positivista en la psicología se han desarrollado cuando menos dos caminos en la investigación científica: la investigación basada en definición de variables, planteamiento de hipótesis y análisis estadístico de los datos empíricos y la de manipulación.

Para los epistemólogos y psicólogos partidarios de la investigación cualitativa en psicología, la epistemología positivista ha sido un recurso para producir conocimiento a través de una simplificación del objeto en

variables, las cuales se convierten en productos del conocimiento por su correlación con otras variables.

Para González Rey (2000) "...la ciencia no es sólo racionalidad, es subjetividad en todo lo que este término implica, es emoción, individuación, contradicción, en fin es expresión íntegra del flujo de la vida humana, que se realiza a través de sujetos individuales, en los que su experiencia se concreta en la forma individualizada de su producción. La representación de la ciencia como actividad supraindividual, que supone la no participación del investigador y el control de su subjetividad, soslaya el carácter interactivo y subjetivo de nuestro objeto, el cual es condición de su expresión comprometida". (p 18). Los postulados de la nueva epistemología de la subjetividad dan cabida a la expresión en términos de probabilidades *a priori* de las creencias del tomador de decisiones que se expresan en una *a priori* subjetiva.

Según González Rey (2000), la investigación cualitativa en psicología desarrolla nuevas formas de investigación que no tienen necesariamente los objetivos de control, predicción y descripción de la investigación positivista; si bien un control estricto de variables no es posible, dada la naturaleza del objeto de estudio, desde la óptica

neopositivista de la autora, la predicción es deseable y la descripción indispensable. Así, en la estimación bayesiana de calificaciones con una *a priori* subjetiva, el fin es predecir acertadamente el fenómeno y en función de esto tomar decisiones que mejoren las decisiones didácticas y de aprovechamiento de los estudiantes.

Cuando una predicción intuitiva o racional (por definición subjetiva) resulta aproximarse también a los fenómenos? (que se nos presentan de la realidad). La realidad está allí y dice Kant que es inaccesible, sin embargo, más allá del positivismo, todo procedimiento que nos permita describirla y predecirla es válido. Desde la óptica de un positivismo ortodoxo, la única *a priori* válida sería la objetiva que estaría basada en observaciones, desde la de la investigación cualitativa, las consideraciones subjetivas también lo son.

Con referencia a la aplicación del teorema de Bayes en la estimación de calificaciones. Consideraremos como *a priori* subjetiva a aquella que es fruto de la intuición del tomador de decisiones o del razonamiento sobre cómo considera que se desarrolla el fenómeno y *a priori* objetiva será aquella que se establece como resultado del análisis estadístico de un conjunto de datos empíricos. Ahora bien, ¿es posible que una *a priori* subjetiva se torne objetiva? Sí, cuando un grupo de

tomadores de decisiones esté de acuerdo en una consideración y ésta sea contrastada con datos empíricos relevantes y se demuestre que existe una correlación positiva significativa, que las funciones son semejantes, o que la predicción es acertada.

Los argumentos a favor de la objetividad basada en la observación sensible de los fenómenos pueden encontrarse en la historia misma de la ciencia, ¿cuándo es posible medir y predecir los fenómenos que se estudian, elaborar teorías coherentes y verificables y dejar a un lado las opiniones personales sobre cómo se desarrollan y a qué se deben los fenómenos? Cuando se desarrollan observaciones, cuando se define claramente qué se está midiendo y cuando se encuentran relaciones funcionales entre los fenómenos que puedan llegar a conformar ecuaciones matemáticas que los describan.

La frecuente falta de confiabilidad del calificador debida al ingrediente de mayor subjetividad en las pruebas de ensayo, ha provocado que las calificaciones no sean comparables sobre una base confiable y sólida, ¿Qué se puede decir también de los exámenes orales que privaron desde el siglo XII antes de Cristo, en China, en la antigua Grecia, en Roma y en el Medievo, y hasta principios del siglo XXI? El

nerviosismo del examinado ha podido influir negativamente en todos los casos.

Sánchez (2001), opina que debe integrarse un paradigma alternativo que incluya la evaluación cualitativa ya que la objetividad, entendida como observación basada en los sentidos, de la evaluación, siempre es relativa y la objetividad en sí misma no puede considerarse el objetivo prioritario de la evaluación.

¿Cuál sería entonces ese objetivo primordial? Comprender la eficiencia y la eficacia de un proceso y tomar decisiones razonables en pro del futuro aprovechamiento de un estudiante y, en el caso de la asignación de calificaciones, emitir un juicio valorativo que infiera el futuro desempeño del alumno, lo que constituiría en sí mismo una indicación de cómo mejorar el proceso de enseñanza-aprendizaje.

En este sentido Landsheere, (1973), afirma que debe desaparecer la época de antagonismo entre los partidarios de la psicometría y los educadores. Las épocas en que un solo examen que se desarrolla en pocas horas o incluso minutos puede decidir sobre una carrera escolar o profesional, no es posible que continúen. Los psicómetras deben humanizar sus números y los educadores introducir más rigor en sus

procedimientos. Con respecto a la opinión de Landsheere, es posible considerar que la cuantificación de las valoraciones subjetivas (basadas en la intuición o en la razón) de los profesores, que se utilizan en el marco de evaluación bayesiano, son una forma de “humanizar” a los psicómetras e introducir elementos cualitativos importantes para la evaluación de los rendimientos de los alumnos sobre una base confiable, ya que la exactitud de las valoraciones subjetivas deberá ser corroborada con medidas futuras de rendimiento.

No es posible olvidar las conclusiones de Theall (1997), Director del Centro de Enseñanza y Aprendizaje de la Universidad de Illinois en Springfield, que afirma que la combinación de datos inexactos aunada a evaluadores inexpertos utilizando el Teorema de Bayes como marco de valoración, produce resultados catastróficos y el olvido de los conceptos de confiabilidad y validez en la aplicación de las pruebas psicológicas. Asimismo, Thompson, (1988) concluye en su investigación que los estudios empíricos han demostrado que las personas que no están familiarizadas con la probabilidad condicional interpretan resultados erróneamente aunque la situación sea simple. Tomando en cuenta las observaciones de Theall y Thompson, los investigadores que utilizan valoraciones subjetivas, deben tener sumo cuidado al introducir datos cualitativos a través de valoraciones y creencias del profesor.

Definitivamente no es posible ignorar la elaboración de pruebas confiables y válidas, ya sea referidas a la norma o a un criterio; pero sobre todo, no debe olvidarse que es imprescindible verificar con datos de ejecuciones relacionadas, el futuro desempeño de los alumnos tomando como base una *a priori* subjetiva. Únicamente la exactitud de la predicción puede ser el elemento de juicio definitivo para que un evaluador la siga utilizando o no.

Stanley y Hopkins, (1972) por su parte, afirman que la evaluación generalmente involucra más subjetividad que la medición, que también la incluye. Así, una evaluación es más subjetiva si depende en mayor medida de quién la está realizando en lugar de depender de qué se está midiendo o evaluando. Tanto la medición como la evaluación involucran aspectos subjetivos en los juicios realizados por maestros y administradores. No es posible caer en la trampa de decir que únicamente se deben tomar en cuenta las puntuaciones de las pruebas o los juicios de los profesores. No es posible eliminar ni las puntuaciones, ni los juicios de expertos. La información proporcionada por las dos fuentes es complementaria.

En la asignación bayesiana de calificaciones, en su modalidad de *a priori* objetiva, lo que se considera como medición es el diseño y construcción de las pruebas, que involucra la especificación de objetivos de aprendizaje en términos de conductas observables y con criterios de ejecución aceptable, la administración y la asignación de puntajes a las pruebas.

En el caso de una *a priori* subjetiva, el profesor asigna números, también a determinadas percepciones suyas del comportamiento del alumno y esta percepción subjetiva cuantificada forma parte de la calificación final. El teorema de Bayes permite trabajar también con las creencias del profesor, además de con las mediciones.

En evaluación educativa, específicamente en evaluación referida a un criterio, los trabajos de Jackson, (1975), Hambleton, (1975), Rouanet, Dominique y Pelnard-Considère, (1975); Wilmick, y Nevels, (1982); constituyen valiosas aplicaciones de la inferencia bayesiana a diferentes ejecuciones con fines de evaluación referida a un criterio.

Curda, (2001), realiza una investigación sobre la utilización de formatos de evaluación formativa para ser usados en cursos por

internet en la maestría de Tecnología Instruccional de la Universidad de Florida y reporta que los resultados fueron satisfactorios.

En medicina, Maduako, Cramp, y Carson, (2002) utilizan marcos de inferencia bayesiana utilizando formatos que involucran datos objetivos para la clasificación y el manejo del carcinoma de cabeza y cuello, lo que puede considerarse como un tipo de evaluación diagnóstica y formativa. El diseño de formatos que sistematicen las observaciones objetivas y que proporcionen sugerencias sobre cómo manejar la subjetividad y asignar probabilidades a la intuición y creencias del tomador de decisiones constituyen temas promisorios para investigaciones futuras en evaluación educativa realizada dentro de un marco bayesiano.

Hoiting, (1988), proporciona un análisis sobre los métodos que combinan probabilidades y concluye que el que proporciona mayor coherencia y exactitud es el bayesiano. En el presente estudio se espera demostrar, al incluir el procedimiento clásico de estimación de probabilidades, precisamente la exactitud en la inferencia del método bayesiano propuesto.

Zapata, y Greer, (2002), se encuentran desarrollando dos proyectos de tesis de doctorado en los que utilizan redes de creencias bayesianas para representar gráficamente, con la ayuda de la computadora, diferentes características del estudiante. Es una forma de que los profesores se enteren, por ejemplo, de los progresos, temas difíciles, y dificultades de los estudiantes. El sistema se denomina VisMod ya que visualiza gráficamente los intereses y las necesidades que deberían ser tomados en cuenta para el aprendizaje de un tópico específico, lo que constituye un enfoque novedoso para la valoración de elementos que generalmente no son tomados en cuenta.

Vanlehn, (1997), propone un procedimiento de evaluación formativa denominado Olae en el que se recopila información sobre el desempeño de los alumnos al resolver problemas tradicionales de física Newtoniana en un programa a ser usado por computadora. Se diseñan formas de realizar valoraciones utilizando el marco bayesiano para la toma de decisiones ya que los procedimientos estadísticos desarrollados con anterioridad se basaban en pruebas de opción múltiple y a juicio de Vanlehn, deterioraban la calidad de la ejecución de los alumnos.

En este estudio, la enseñanza de la Físicoquímica también se realiza resolviendo problemas de aplicación y aunque existe una

tendencia a utilizar exámenes de opción múltiple con evaluación referida a la norma, (como el que se desarrolló en el CENEVAL para el examen de egreso a la licenciatura, particularmente para el caso de la carrera de Ingeniería Química Industrial), sería recomendable que se utilizaran los exámenes basados en resolución de problemas con valoraciones definidas para el procedimiento y la exactitud de la respuesta dentro de un marco bayesiano, ya que se coincide con el juicio de Van Lehn en el sentido de que los exámenes de opción múltiple deterioran la calidad de las ejecuciones en una gran proporción de los casos.

Vanlehn, y Zhendong, (1999), desarrollaron un sistema tutorial para física llamado Andes que moldea al estudiante utilizando una red bayesiana. Estudiaron los efectos de variar los parámetros numéricos del modelador en lugar de tomar en cuenta el número de entradas equivocadas y contar el número de errores cometidos, como es la práctica común. Se evaluó como probabilidad *a priori* si el sistema tutorial utilizaba *feedback* o si insistía en que el estudiante corrigiera sus errores. Los resultados demostraron que la existencia del *feedback* mejoraba la eficiencia del asesor. El procedimiento bayesiano fue útil para asignar el crédito al problema y desecharlo. Como puede observarse, el tipo de las valoraciones de un programa de enseñanza

puede diversificarse y será la constatación de su eficacia el criterio más adecuado para juzgar su inclusión dentro de un procedimiento de evaluación formativa como el desarrollado por Vanlehn y Zendog.

Koedinger, (1997), propuso un sistema inteligente tutorial por computadora para la enseñanza de las matemáticas en una escuela secundaria urbana. El enfoque consiste en un diseño centrado en el cliente que se desarrolló para lograr la conducta experta en los objetivos curriculares. Se utiliza un procedimiento de estimación bayesiana que identifica las fortalezas y debilidades relacionadas con la producción de reglas de acuerdo con el modelo cognitivo utilizado. La *a priori* utilizada fue objetiva, sin embargo, sería factible utilizar un conjunto de probabilidades *a priori* subjetivas si se diseñara un conjunto de valoraciones que a juicio del profesor fueran relevantes, que podrían obtenerse al aplicar un cuestionario específico, también dentro del programa experto.

Mislevy, Steinberg, y Almond, (1999) desarrollaron formas novedosas alternativas de realizar valoraciones, a través de las cuales puede evidenciarse el aprendizaje en el moldeamiento. Utilizan redes de inferencia bayesiana para manejar el conocimiento incierto acerca del desempeño del estudiante y tomar decisiones en diferentes situaciones.

Cuando el proceso de moldeamiento aún no había iniciado, se utilizaban distribuciones no informativas como punto de partida. En el caso de la estimación de calificaciones, precisamente, una de las bondades del marco bayesiano es la posibilidad de ir aumentando la precisión de la predicción realizada, con referencia a un fenómeno dado, a través de la realización de nuevas observaciones. Es posible iniciar con una *a priori* no informativa asignando probabilidades iguales a los posibles resultados, y a medida que los datos se recaben, es posible incluirlos y lograr inferencias más exactas.

También, el estudio de Rosas Mercado (1984) sobre evaluación referida a un criterio, en el que con base en registros de calificaciones anteriores de una materia calificada de acuerdo al grado de dominio presentado por el aprendiz en relación con los objetivos de la materia y la calificación de un primer examen, se estimaron, las calificaciones finales más probables; constituyen antecedentes al desarrollo del presente estudio.

### ***Planteamiento del problema***

Como puede observarse, las investigaciones de Jackson, (1975), Hambleton, (1975), Rouanet, Dominique y Pelnard Considère, (1975),

Wilmick, y Nevels, (1982), Curda, (2001), Koedinger, (1997), Mislevy, Steiberg, y Almond, (1999) y Maduako, Cramp, y Carson, (2001), Vanlehn, y Zhendong, y el de Rosas Mercado (1984), responden a la interpretación de la inferencia bayesiana de Jeffreys sobre la probabilidad *a priori* objetiva, ya que utilizan datos objetivos recabados para la estimación de probabilidad. Por otra parte, los estudios de Zapata, y Greer, que se están desarrollando en este momento como tesis doctoral responden a la interpretación subjetiva del Teorema de Bayes. Lo que lleva a plantear lo siguiente: ¿es aplicable la tesis de Savage sobre probabilidades *a priori* subjetivas a la estimación bayesiana de calificaciones? Y si esto es cierto, ¿en qué condiciones sería recomendable su uso?

¿Y qué de los casos en los que el maestro no cuenta con datos objetivos de ninguna especie? ¿Es posible también en esta situación aplicar el teorema de Bayes y obtener resultados satisfactorios? El presente estudio pretende resolver estas cuestiones proponiendo dos nuevos procedimientos estadísticos basados en la inferencia bayesiana para la predicción de calificaciones o notas, con fines de evaluación formativa, en tres situaciones diferentes:

1. *A priori* objetiva. (Réplica del estudio anteriormente citado de Rosas Mercado (1984)). Situación que puede derivarse de dos estados diferentes: en primera instancia, el conocimiento de registros de calificaciones anteriores de los alumnos que han tomado la materia con ese profesor y bajo los mismos criterios de medición y evaluación de las ejecuciones; y en segunda instancia, el conocimiento de los registros de calificaciones de los alumnos en otras materias.

2. *A priori* no informativa. Situación derivada del hecho de que el profesor no cuente con registros de calificaciones anteriores, ni siquiera la calificación de una prueba.

3. *A priori* subjetiva. Situación en que el profesor hace uso de apreciaciones subjetivas sobre el desempeño de sus alumnos, pero no cuenta ni con registros de calificaciones anteriores, ni con resultados de la aplicación de una prueba.

### **Propósito y fundamentación**

El propósito de esta tesis es responder a la pregunta: ¿Qué procedimiento bayesiano constituye un mejor método de estimación de la calificación definitiva en función de diferentes grados de incertidumbre: 1) Registros objetivos de calificaciones de un examen evaluado en función de los criterios especificados en los objetivos de la materia. (Menor grado de incertidumbre) 2) Apreciaciones subjetivas

de la ejecución futura de un alumno. (Mayor grado de incertidumbre), o 3) el que utiliza principios de probabilidad clásica como base de cálculo sin contar con registros de calificaciones. (El mayor grado de incertidumbre) ? Por mejor método se entiende aquél que permita establecer un mayor grado de concordancia con los eventos relacionados ocurridos después de que la predicción ha tenido efecto. Esto es, el mejor método será el que prediga más acertadamente.

Los posibles usos de los procedimientos delineados serán en primera instancia dos nuevas formas de estimar calificaciones ante diferentes tipos de incertidumbre y la posibilidad de contar con la predicción del futuro desempeño del alumno al principio o apenas iniciado un curso, lo que posibilitará diferentes acciones correctivas que pudieran disminuir las probabilidades de que el estudiante repruebe el curso con el consiguiente desperdicio de recursos.

Los tres métodos de estimación de calificaciones propuestos se realizarán de acuerdo con los principios de la estadística bayesiana. Los resultados de cada predicción serán contrastados utilizando una prueba de hipótesis apropiada para el tipo de datos obtenidos y se discutirán los resultados.

Con el fin de no desviar la atención de la exposición, se anexará un glosario en el que se definen los principales términos utilizados.

## Capítulo II

### Método

Se proponen tres procedimientos estadísticos; basados en el Teorema de Bayes para la predicción de calificaciones o notas en un tipo de evaluación referida a un criterio. Los procedimientos propuestos pretenden predecir la calificación final obtenida por el alumno, ante diferentes tipos de incertidumbre, relacionados con la cantidad de información que el profesor o tomador de decisiones posea en el momento de tomar la decisión. El procedimiento con menor incertidumbre es el que se plantea con base en el análisis de los resultados observados de una prueba evaluada con referencia a un criterio. El segundo procedimiento utiliza los principios de probabilidad clásica, sin información disponible, para calcular la *a priori*, por lo que es de suponer que involucra el mayor grado de incertidumbre. Y, el tercer procedimiento, basado también en la estadística bayesiana, involucra una incertidumbre media, ya que cuantifica las creencias subjetivas del tomador de decisiones.

## Participantes

De una población de 514 estudiantes de un solo profesor de la materia de Fisicoquímica que se imparte actualmente en el tercer semestre de la carrera de Ingeniería Química Industrial se eligió una muestra representativa. Para determinar el tamaño de la muestra se utilizó el procedimiento de muestreo por proporciones descrito por Cochran (1971).

## Variables

### Dependientes

*A posteriori* objetiva. La estimación de probabilidad computada como resultado de la aplicación del procedimiento desarrollado para la *a priori* objetiva.

*A posteriori* no informativa. Esto es, la inferencia resultante de la aplicación del procedimiento relacionado con la *a priori* no informativa.

*A posteriori* subjetiva. Es decir, la probabilidad calculada como resultado del procedimiento desarrollado para la *a priori* subjetiva.

## Independientes

*A priori* objetiva. Procedimiento bayesiano consistente en considerar, como base de cálculo, la probabilidad condicional de las frecuencias de calificaciones de un examen inicial y las frecuencias de calificaciones finales correspondientes que aparecen en la muestra.

*A priori* no informativa. Procedimiento bayesiano consistente en considerar como base de cálculo la existencia de probabilidades iguales para cada una de las calificaciones que aparecen en la muestra.

*A priori* subjetiva. Procedimiento bayesiano consistente en considerar como base de cálculo la consideración subjetiva del evaluador, que en este caso fue que las probabilidades previas se distribuían normalmente y que fijando “intuitivamente” la media, se asignaran las probabilidades correspondientes a cada categoría de

acuerdo con la probabilidad asociada bajo la superficie de la curva normal.

## Diseño

Se aplicaron tres procedimientos de inferencia bayesiana a la misma muestra de sujetos y las inferencias resultantes se compararon con los datos de la población para verificar cuál de las predicciones fue más certera. Lo cual puede expresarse esquemáticamente en la figura 1:

n	$X_1$	$Y_1$	N
n	$X_2$	$Y_2$	N
n	$X_3$	$Y_3$	N

Figura 1. Representación esquemática del diseño de investigación de los efectos de diferentes tipos de *a priori* en la estimación bayesiana de calificaciones. En la cual  $n$  representa la muestra de participantes,  $X_1$  representa a la *a priori* objetiva,  $X_2$  representa a la *a priori* no informativa y  $X_3$  representa la *a priori* subjetiva. Donde  $Y_1$  representa la probabilidad *a posteriori* estimada a partir del procedimiento para la *a priori* objetiva,  $Y_2$  representa la probabilidad *a posteriori* estimada a partir del procedimiento para la *a priori* no informativa y  $Y_3$  la probabilidad *a posteriori* estimada a partir del procedimiento para la *a priori* subjetiva y donde  $N$  representa los datos de la población.

## Hipótesis

### **Nula**

El procedimiento de asignación de calificaciones para una *a priori* objetiva es igualmente válido que el procedimiento de asignación de calificaciones para una *a priori* no informativa y el de una *a priori* subjetiva, para predecir con exactitud el futuro desempeño del estudiante.

Por procedimiento válido se entiende aquél cuya inferencia resultante prediga con mayor exactitud los datos de la población. En consecuencia, el mejor procedimiento será aquél cuya inferencia resultante se desvíe en menor cuantía de los datos emanados de la población.

### **Alternativa**

El procedimiento de estimación de calificaciones para una *a priori* objetiva es el que presenta una validez predictiva mayor que el procedimiento de estimación de calificaciones para una *a priori* no informativa y que el de una *a priori* subjetiva, para predecir con exactitud el futuro desempeño del estudiante.

## Procedimientos

### *A priori objetiva*

A continuación se resumirá el procedimiento a utilizar (Rosas Mercado, 1984): 1) Elección de una muestra aleatoria. 2) Determinar, para la muestra, las frecuencias de calificaciones iniciales  $C_I$  y las definitivas  $C_D$ . 3) Determinar, para la población, las frecuencias de calificaciones iniciales  $C_I$  y las definitivas  $C_D$ . 4) Cálculo de las frecuencias esperadas de la población sobre la base de la  $N$  de la muestra para permitir la comparación. 5) Estimación de las probabilidades *a priori* de la muestra. Se calculan las probabilidades de obtener cada una de las calificaciones definitivas  $P(C_{D_i})$  y éstas son tomadas como las probabilidades *a priori* objetivas. 6) Estimación de las probabilidades de intersección de la muestra.

$P(C_I \cap C_D)$  según:

$$P(C_I \cap C_D) = P(C_I | C_D) P(C_D) \quad 2.1$$

mismas que servirán de base para el cálculo de las probabilidades *a*

*posteriori*:

$P(C_D | C_I)$  de acuerdo al Teorema de Bayes según:

$$P(C_D | C_I) = P(C_I \cap C_D) / P(C_I)_M$$

donde  $P(C_i)_M$  es la probabilidad marginal de  $C_i$ . Lo anterior implica la aplicación del Teorema de Bayes, ya que:

$$P(C_{Dk} | C_i) = \frac{P(C_{Dk})P(C_i \setminus C_{Dk})}{\sum_{i=1}^n P(C_{Di})P(C_i \setminus C_{Di})} \quad 2.2$$

$$P(C_{Dk} | C_i) = \frac{P(C_{Dk} \cap C_i)}{\sum_{i=1}^n P(C_{Di} \cap C_i)} \quad 2.3 a$$

$$P(C_{Dk} | C_i) = \frac{P(C_{Dk} \cap C_i)}{P(C_i)_M} \quad 2.3 b$$

Resulta entonces claro que la probabilidad marginal de  $C_i$  está dada por:

$$P(C_i)_M = \sum_{i=1}^n P(C_{Di} \cap C_i) \quad 2.3 c$$

es decir, es la suma de probabilidades de las intersecciones de las diferentes  $C_{Di}$  con una calificación  $C_i$  dada. 7) Estimación del valor esperado. De acuerdo con los principios de valor esperado expresados por Ya Lun Chou (1977), se asigna un número a cada posible calificación y se determina, posteriormente, la calificación definitiva esperada, para cada calificación del primer examen parcial.

NA	S	B/MB
1	2	3

$$C_D^* = (1)P(NA_D) + (2)P(S_D) + (3)P(B_D MB_D) \quad 2.4$$

8) Determinar la calificación definitiva esperada, para cada calificación del primer examen parcial:

$$C_D^{**} = (1)P(NA_D \setminus C_I) + (2)P(S_D \setminus C_I) + (3)P(BoMB_D \setminus C_I) \quad 2.5$$

9) Establecer en cuántos casos de la población concuerdan los valores de  $C_D$ , con los valores esperados  $C_D^*$  y  $C_D^{**}$ , para cada calificación  $C_I$ .

### ***Elección de una muestra aleatoria.***

La población de la cual se tomó la muestra fue de 514 estudiantes. De acuerdo con los postulados del Teorema del Límite Central, que establece que cuando se dan  $n$  variables aleatorias independientes que tienen la misma distribución, cualesquiera que ésta sea, las medias y varianzas de la población tienen distribuciones similares a las medias y varianzas de la muestra.

El procedimiento elegido fue el del muestreo por proporciones debido a la sencillez del mismo y a que los parámetros necesarios son fácilmente obtenidos de una muestra previa de la población. Los eventos individuales se clasificaron en dos clases: C y C', donde C representa al subconjunto de los alumnos aprobados y C' a su complemento, los reprobados. El margen de error  $d$  considerado como aceptable, en la proporción estimada  $P$ , de unidades pertenecientes a la clase C y un riesgo  $\alpha$  de que el error real sea mayor que  $d$ , se fijaron en los siguientes valores:

$$d = 0.10$$

$$\alpha = 0.05$$

Esta muestra fue tomada de acuerdo a los principios del muestreo por proporciones con una desviación permitida como aceptable del 0.05%. Los datos de un muestreo aleatorio previo revelaron una proporción del 36% para los no aprobados y de 64% para los aprobados. (Cochran, 1971).

Dada la fórmula para la determinación del tamaño de la muestra por proporciones:

$$n = \frac{\left[ \frac{t^2(p \cdot q)}{d^2} \right]}{1 + \frac{1}{N} \left[ \left( \frac{t^2(p \cdot q)}{d^2} \right) - 1 \right]}$$

Donde:

p = proporción de casos aprobados en la muestra previa.

q = proporción de casos en la premuestra.

t = valor de desvío normal (desviación estándar)

correspondiente al nivel  $\alpha$  de confianza deseado.

d = desviación permitida como aceptable en la predicción de la calificación.

Si  $p = 0.75$ ,  $q = 0.25$ ,  $t = 1.96$  ( $\alpha = 0.05$ ),  $d = 0.05$  como  $N = 514$

$$n = \frac{\left[ \frac{(1.96)^2(0.75 \times 0.25)}{(0.05)^2} \right]}{1 + \frac{1}{514} \left[ \left( \frac{(1.96)^2(0.75 \times 0.25)}{(0.05)^2} \right) - 1 \right]} = 184.86$$

Los participantes fueron 185 alumnos de un total de 514 de la materia de Fisicoquímica III, de la carrera de Ingeniería Química Industrial de la Escuela Superior de Ingeniería Química del Instituto Politécnico Nacional.

***Determinación de las frecuencias de calificaciones iniciales y definitivas de la muestra.***

La asignación de calificaciones fue realizada en relación con la perfección. Esto es, el puntaje estaba relacionado ordinalmente con el grado de dominio demostrado por los sujetos en concordancia con los objetivos de la materia, redactados estos conductualmente.

Los puntajes estaban expresados en forma de porcentajes y después fueron transformados a letras de acuerdo con las siguientes reglas: de 0 - 59% - NA; de 60 -84% - S; de 85 - 94% B y de 95 -100 % MB.

***Tabla 1***

***Calificaciones iniciales y definitivas de la población***

Calificación inicial ( $C_i$ )	Calificación definitiva ( $C_D$ )			TOTAL
	$NA_D$	$S_D$	$B_D, MB_D$	
$NA_i$	84	60	23	167
$S_i$	53	61	45	159
$B_i, MB_i$	13	29	146	188
TOTAL	150	150	214	514

**Tabla 2*****Frecuencias esperadas (Población)***

Calificación inicial ( $C_i$ )	Calificación definitiva ( $C_D$ )		
	$NA_D$	$S_D$	$B_D, MB_D$
$NA_i$	30.2335	21.5953	8.2782
$S_i$	19.0759	21.9553	16.1965
$B_i, MB_i$	4.6790	10.4377	52.5486
TOTAL	53.9884	53.9883	77.0233

Los cálculos de las frecuencias esperadas se realizaron con base a una  $n$  de 185 para permitir posteriormente la comparación con la prueba de hipótesis y de acuerdo con la siguiente fórmula:

$$n(C_i \cap C_D) / n(P) \cdot N(M)$$

donde  $n(P)$  es igual al número de observaciones de la población y  $N(M)$  al de la muestra.

**Tabla 3**

***Elementos de la muestra y frecuencias observadas según calificación inicial ( $C_I$ ) y calificación definitiva ( $C_D$ ).***

Calificación inicial ( $C_I$ )	Calificación definitiva ( $C_D$ )			TOTAL
	NA	S	B,MB	
NA	36	20	8	64
S	19	23	17	59
B, MB	2	11	49	62
TOTAL	57	54	74	185

***Determinación de las probabilidades a priori objetivas.***

A partir de las frecuencias de calificaciones definitivas de la muestra, se calcularon las probabilidades correspondientes  $P(C_{Di})$

$$P(NA_D) = n(NA_D) / n(S) = 57 / 185 = 0.3081$$

$$P(S_D) = n(S_D) / n(S) = 54 / 185 = 0.2919$$

$$P(B_D) = n(B_D) / n(S) = 71 / 185 = 0.3838$$

$$P(MB_D) = n(MB_D) / n(S) = 3 / 185 = 0.0162$$

Las probabilidades anteriores fueron calculadas de acuerdo con el principio de razón insuficiente de Bernoulli, de la probabilidad clásica, y sirvieron de base para el cálculo de la *a priori* objetiva y la *a priori* no informativa. Dado que la probabilidad de calificación definitiva de MB es muy pequeña, se optó por unirla con la categoría anterior: B; por lo que:

$$P(B_D MB_D) = 0.4000$$

*Estimación de probabilidades de intersección de las calificaciones de la muestra.*

El siguiente paso es la determinación de las probabilidades de intersección  $P(C_I \cap C_D)$  de acuerdo con:

$$P(C_I \cap C_D) = P(C_I | C_D) * P(C_D) \text{ y} \quad 2.1$$

$$P(C_I | C_D) = n(C_I \cap C_D) / n(C_D) * P(C_D) \quad (2)$$

Por lo que según los datos de los elementos de la muestra con base en diferentes calificaciones iniciales y definitivas (Tabla 3) y de las probabilidades de calificación definitiva calculadas de acuerdo con los principios clásicos ( $PC_{Di}$ ), se tiene que:

$$P(NA_i \cap NA_D) = (36 / 57) * 0.3081 = 0.1946$$

$$P(NA_i \cap S_D) = (20 / 54) * 0.2919 = 0.1081$$

$$P(NA_i \cap B_D MB_D) = (8 / 74) * 0.4000 = 0.0432$$

$$P(S_i \cap NA_D) = (19 / 57) * 0.3081 = 0.1027$$

$$P(S_i \cap S_D) = (23 / 54) * 0.2919 = 0.1243$$

$$P(S_i \cap B_D MB_D) = (17 / 74) * 0.4000 = 0.0919$$

$$P(B_i MB_i \cap NA_D) = (2 / 57) * 0.3081 = 0.0108$$

$$P(B_i MB_i \cap S_D) = (11 / 54) * 0.2919 = 0.0595$$

$$P(B_i MB_i \cap B_D MB_D) = (49/54) * 0.4000 = 0.2649$$

Cálculo de las probabilidades marginales:

$$P(C_i)_M = \sum P(C_i \cap C_{Di})$$

$$P(NA_D)_M = 0.1946 + 0.1081 + 0.0432 = 0.3459$$

$$P(S)_M = 0.1207 + 0.1243 + 0.0919 = 0.3189$$

$$P(B_D MB_D)_M = 0.0108 + 0.0595 + 0.2649 = 0.3352$$

Los resultados se resumen en la tabla 4.

**Tabla 4**

***Probabilidades de intersección de las calificaciones de la muestra***

Calificación inicial ( $C_I$ )	Calificación definitiva ( $C_D$ )			
	$NA_D$	$S_D$	$B_D MB_D$	$P(C_I)_M$
$NA_I$	0.1946	0.1081	0.0432	0.3459
$S_I$	0.1027	0.1243	0.0919	0.3189
$B_I, MB_I$	0.0108	0.0595	0.2649	0.3352

***Estimación de las probabilidades a posteriori con base en una a priori objetiva.***

El cálculo de las probabilidades *a posteriori*, según el teorema de Bayes, se basa en la utilización de las probabilidades marginales de acuerdo con el procedimiento de Rosas Mercado (1984):

$$P(C_{Dk} | C_I) = \frac{P(C_{Dk} \cap C_I)}{P(C_I)_M} \quad 2.3 b$$

Tomando en cuenta que:

$$P(C_{Dk} | C_I) = \frac{P(C_{Dk} \cap C_I)}{P(C_I)_M} \quad 2.3b$$

y la ecuación para calcular las intersecciones de las calificaciones iniciales con las definitivas:

$$P(C_{Di} \cap C_I) = P(C_{Di}) P(C_I | C_{Di}) \quad 2.1$$

entonces, la ecuación también puede tomar la forma:

$$P(C_{Dk} | C_I) = P(C_{Di}) \frac{P(C_{Dk} \cap C_I)}{P(C_I)_M} \quad 2.4$$

que constituye la forma clásica del Teorema de Bayes, en donde:  $P(C_{Di})$  es la probabilidad *a priori* y expresa el estado del conocimiento antes de que se conozcan las calificaciones iniciales. El primer término de la ecuación es la probabilidad *a posteriori* y expresa el estado del conocimiento sobre las calificaciones definitivas después de realizar ciertas consideraciones. Las ecuaciones 2.3b y 2.4 son equivalentes.

Donde la probabilidad marginal de  $C_I$  está dada por:

$$P(C_I)_M = \sum_{i=1}^n P(C_{Di} \cap C_I) \quad 2.3c$$

Así, de acuerdo con la ecuación (2.3b) fueron calculadas las probabilidades *a posteriori*, a continuación se detalla el cálculo y los resultados se resumen en la tabla 5.

$$P(N_{A_D} | NA_i) = 0.1946/0.3459 = 0.5626$$

$$P(S_D | NA_i) = 0.1081/0.3459 = 0.3125$$

$$P(B_D MB_D | NA_i) = 0.0432/ 0.3459 = 0.1249$$

$$P(N_{A_D} | S_i) = 0.1027/ 0.3189 = 0.3220$$

$$P(S_D | S_i) = 0.1243/ 0.3189 = 0.3898$$

$$P(B_D MB_D | S_i) = 0.0919/0.3189 =0.2882$$

$$P(N_{A_D} | B_i MB_i) = 0.0108/0.3352 =0.0322$$

$$P(S_D | B_i MB_i) = 0.0595/ 0.3352 = 0.1775$$

$$P(B_D MB_D | B_i MB_i) =0.2649/ 0.3552 = 0.7903$$

**Tabla 5**  
**Probabilidades *a posteriori* a partir de una *a priori***  
**objetiva**

Calificación	Probabilidades <i>a posteriori</i>		
	$P(C_D   NA_i)$	$P(C_D   S_i)$	$P(C_D   B_i MB_i)$
Definitiva			
$A_D$	0.5626	0.3220	0.0322
$S_D$	0.3125	0.3898	0.1775
$B_D, MB_D$	0.1249	0.2882	0.7903

**Estimación del valor esperado.**

Se realizaron los cálculos pertinentes para encontrar el valor esperado más probable para una *a priori* objetiva y posteriormente la calificación definitiva esperada de acuerdo con la siguiente expresión:

$$C_D^* = (1)P(NA_D) + (2)P(S_D) + (3)P(B_D MB_D) \quad 2.4$$

$$C_{NAD}^* = 1X0.5626 + 2X0.3125 + 3X0.01249 = 1.5623 \quad (S)$$

$$C_{SD}^* = 1X0.3220 + 2X0.3898 + 3X0.2882 = 1.9662 \quad (S)$$

$$C_{BoMBD}^* = 1X0.0322 + 2X0.1775 + 3X0.7903 = 2.7584 \quad (B \text{ o } MB)$$

***A priori no informativa***

Para la *a priori* no informativa se procedió asignar probabilidades iguales para cada calificación definitiva sin importar la calificación inicial obtenida, se aplicó el principio de Bernoulli.

**Estimación de las probabilidades de intersección.**

Se procedió a calcularlas según:

$$P(C_I \cap C_D) = P(C_I | C_D) P(C_D)$$

las probabilidades obtenidas se resumen en la tabla 6 y son aquéllas que servirán de base para el cálculo de las probabilidades *a posteriori*.

**Tabla 6**

***Probabilidades de intersección de las calificaciones para una a priori no informativa***

Calificación inicial ( $C_I$ )	Calificación definitiva ( $C_D$ )			$P(C_I)_M$
	$NA_D$	$S_D$	$B_D, MB_D$	
$NA_I$	0.1111	0.1111	0.1111	0.3333
$S_I$	0.1111	0.1111	0.1111	0.3333
$B_I, MB_I$	0.1111	0.1111	0.1111	0.3333

***Cálculo de las probabilidades a posteriori.***

Se procedió a realizar la estimación de las probabilidades *a posteriori* según:

$P(C_D | C_I)$  de acuerdo con Teorema de Bayes según:

$$P(C_D | C_I) = P(C_I \cap C_D) / P(C_I)_M$$

donde  $P(C_i)_M$  es la probabilidad marginal de  $C_i$ .

Los resultados se resumen en la tabla 7.

**Tabla 7**

***Probabilidades a posteriori a partir de una a priori no informativa***

Calificación	Probabilidades a posteriori		
	$P(C_D   NA_t)$	$P(C_D   S_t)$	$P(C_D   B_tMB_t)$
Definitiva			
$NA_D$	20.5535	20.5535	20.5535
$S_D$	20.5535	20.5535	20.5535
$B_D, MB_D$	20.5535	20.5535	20.5535

**Estimación del valor esperado.**

Se efectuaron las operaciones pertinentes para establecer el valor esperado más probable para una *a priori* no informativa y posteriormente la calificación definitiva esperada de acuerdo con la siguiente expresión:

$$C_D^* = (1)P(NA_D) + (2)P(S_D) + (3)P(B_D, MB_D) \quad 2.4$$

$$C_D^* = 1 \times 0.3333 + 2 \times 0.3333 + 3 \times 0.3333 = 2 \quad (S)$$

Resultando igual para todos los casos.

### ***A priori subjetiva***

#### **Estimación de la probabilidad *a priori*.**

El procedimiento que se siguió fue el siguiente: partiendo de una reflexión sobre el problema, esto es, se trata de predecir calificaciones definitivas si se conocen las calificaciones iniciales, se trató de investigar si existía una función de distribución conocida para el problema, la respuesta es: la distribución normal de probabilidad. Partiendo del hecho de que los puntajes originales de los exámenes tenían un rango que iba de 0 - 10, y que una calificación de NA equivale al rango 0 - 5.9, una calificación de S al de 6 - 7, una calificación de B a 8 y 9; y una calificación de MB a 10; se equipara la mediana de la distribución de calificaciones con 0, y se toman las probabilidades correspondientes con la curva normal, para la estimación de las probabilidades *a priori*.

Asignando una probabilidad *a priori* por calificación definitiva:

$$P(\text{NA}_D) = 0.6000;$$

$$P(\text{S}_D) = 0.2400;$$

$$P(\text{B}_D \text{MB}_D) = 0.1600$$

Y con base en ellos, desglosando por calificación inicial, se asignan las siguientes probabilidades de intersección de esta distribución ficticia.

**Estimación de las probabilidades de intersección.**

De acuerdo con:

$$P(C_I \cap C_D) = P(C_I | C_D) P(C_D) \quad 2.1$$

Los resultados se resumen en la tabla 8.

**Tabla 8**

***Probabilidades de intersección de la distribución subjetiva***

Calificación inicial ( $C_I$ )	Calificación definitiva ( $C_D$ )			$P(C_I)_M$
	$NA_D$	$S_D$	$B_D, MB_D$	
$NA_I$	0.37	0.06	0.03	0.46
$S_I$	0.19	0.15	0.05	0.39
$B_I, MB_I$	0.04	0.03	0.08	0.15

**Estimación de las probabilidades *a posteriori***

Según:

$P(C_D | C_I)$  de acuerdo con el Teorema de Bayes según:

$$P(C_D | C_I) = P(C_I \cap C_D) / P(C_I)_M$$

donde  $P(C_I)_M$  es la probabilidad marginal de  $C_I$ .

Calculando las probabilidades *a posteriori*:

$$P(\text{NA}_D | \text{NA}_I) = 0.37/0.46 = 0.8044$$

$$P(\text{S}_D | \text{NA}_I) = 0.06/0.46 = 0.1304$$

$$P(\text{B}_D \text{ MB}_D | \text{NA}_I) = 0.03/0.46 = 0.0652$$

$$P(\text{NA}_D | \text{S}_I) = 0.19/0.39 = 0.4872$$

$$P(\text{S}_D | \text{S}_I) = 0.15/0.39 = 0.3846$$

$$P(\text{B}_D \text{ MB}_D | \text{S}_I) = 0.05/0.39 = 0.1282$$

$$P(\text{NA}_D | \text{B}_I \text{ MB}_I) = 0.04/0.15 = 0.2667$$

$$P(\text{S}_D | \text{B}_I \text{ MB}_I) = 0.03/0.15 = 0.2000$$

$$P(\text{B}_D \text{ MB}_D | \text{B}_I \text{ MB}_I) = 0.08/0.15 = 0.5333$$

Los resultados se resumen en la tabla 9.

**Tabla 9****Probabilidades a posteriori para una a priori subjetiva**

Calificación Inicial	Probabilidades de calificación definitiva		
	$P(NA_D)$	$P(S_D)$	$P(B_D MB_D)$
$NA_I$	0.8044	0.1304	0.0652
$S_I$	0.4872	0.3846	0.1282
$B_I MB_I$	0.2667	0.2000	0.5333

**Estimación del valor esperado.**

De acuerdo con la expresión de valor esperado, se procedió a realizar los cálculos.

$$C_D^* = (1)P(NA_D) + (2)P(S_D) + (3)P(B_D MB_D) \quad 2.4$$

$$C_{NA_D}^* = 1 \times 0.8044 + 2 \times 0.1304 + 3 \times 0.0652 = 1.2608 \quad (NA)$$

$$C_{S_D}^* = 1 \times 0.4872 + 2 \times 0.3846 + 3 \times 0.1282 = 1.6410 \quad (S)$$

$$C_{BoMBD}^* = 1 \times 0.2667 + 2 \times 0.2 + 3 \times 0.5333 = 2.2666 \quad (S)$$

## Capítulo III

### Resultados

#### *Prueba de hipótesis: ( $\chi^2$ ) ji cuadrada*

Con objeto de determinar cuál de los tres métodos bayesianos de estimación de probabilidad: con base en una *a priori* objetiva, no informativa y subjetiva permitió una predicción más precisa de las calificaciones definitivas: se procedió a realizar el siguiente análisis de resultados con base en:

i. La prueba  $\chi^2$  (ji cuadrada). Se eligió debido a que los datos con los que se trabajó fueron ordinales, las probabilidades son fácilmente transformables a frecuencias y la esencia de la prueba consiste en la comparación de frecuencias observadas (las de los tres métodos inferenciales) con frecuencias teóricas o esperadas (los datos reales fr las frecuencias de calificaciones de la población de 514 estudiantes). Así, mientras menor sea el valor de  $\chi^2$  para procedimientos de

estimación de probabilidades *a posteriori*, y dado que se utilizaron las frecuencias de la población como las teóricas o esperadas en la prueba, mayor será el grado de acuerdo de las predicciones con los datos reales de la población.

La hipótesis nula establece que los tres métodos bayesianos de predicción de la calificación final son igualmente válidos: La hipótesis alterna  $H_1$ , que los tres métodos son diferentes.

El nivel de significación elegido fue 0.05. Se considera que la distribución muestral conforme a la hipótesis nula está distribuida como  $\chi^2$  con grados de libertad =  $(k - 1)(r - 1)$ .

La tabla 10 muestra las frecuencias observadas correspondientes con los tres procedimientos bayesianos propuestos y las frecuencias esperadas, que en este caso corresponden con las de la población.

El cálculo se realizó de acuerdo con la siguiente expresión, para el caso de  $r$  clases:

$$\chi^2 = \sum_1^k \sum_1^r (f_1 - F_1)^2 / F_1$$

**Tabla 10**

***Frecuencias observadas y esperadas para los tres procedimientos bayesianos de estimación de calificaciones.***

Frecuencias:

Población:	C <sub>D</sub>			F <sub>ij</sub>
	NA <sub>D</sub>	S <sub>D</sub>	B <sub>D</sub> MB <sub>D</sub>	
NA <sub>I</sub>	30.2335	21.5953	8.2782	
S <sub>I</sub>	19.0759	21.9553	16.1965	
B <sub>I</sub> MB <sub>I</sub>	4.6790	10.4577	52.5486	

Objetiva:	C <sub>D</sub>			F <sub>i/o</sub>
	NA <sub>D</sub>	S <sub>D</sub>	B <sub>D</sub> MB <sub>D</sub>	
NA <sub>I</sub>	36	20	8	
S <sub>I</sub>	19	23	17	
B <sub>I</sub> MB <sub>I</sub>	2	11	49	

Subjetiva:	C <sub>D</sub>			F <sub>i/s</sub>
	NA <sub>D</sub>	S <sub>D</sub>	B <sub>D</sub> MB <sub>D</sub>	
NA <sub>I</sub>	51.4816	8.3456	4.1728	
S <sub>I</sub>	28.7448	22.6914	7.5638	
B <sub>I</sub> MB <sub>I</sub>	16.5354	12.4000	33.0646	

No informativa:	C <sub>D</sub>			F <sub>i/Ni</sub>
	NA <sub>D</sub>	S <sub>D</sub>	B <sub>D</sub> MB <sub>D</sub>	
NA <sub>I</sub>	20.5535	20.5535	20.5535	
S <sub>I</sub>	20.5535	20.5535	20.5535	
B <sub>I</sub> MB <sub>I</sub>	20.5535	20.5535	20.5535	

$$X^2_{O} = 3.1207$$

$$X^2_{S} = 72.2620$$

$$X^2_{NI} = 105.8702$$

El valor de  $\chi^2$  (ji cuadrada) con 4 grados de libertad para los tres procedimientos fue para la a priori objetiva de 3.12, para la subjetiva de 72.26 y para la no informativa de 105.87:

Las comparaciones entre las frecuencias esperadas: los datos de la población y los procedimientos para la *a priori* subjetiva y no informativa resultaron significativos al nivel del 0.05 %. El método basado en la *a priori* objetiva no resultó significativo a ese mismo nivel, sin embargo, es importante aclarar que la predicción que más se acerca a los datos reales de la población es precisamente la basada en la *a priori* objetiva y que se está utilizando la prueba de  $\chi^2$  cuadrada como una prueba de ajuste de curva.

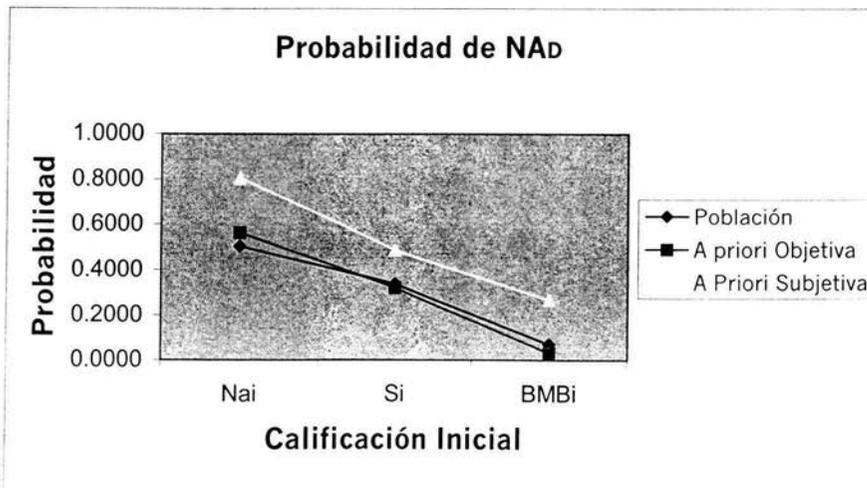


Figura 2. Probabilidad de obtener NA<sub>p</sub> dada la calificación inicial.

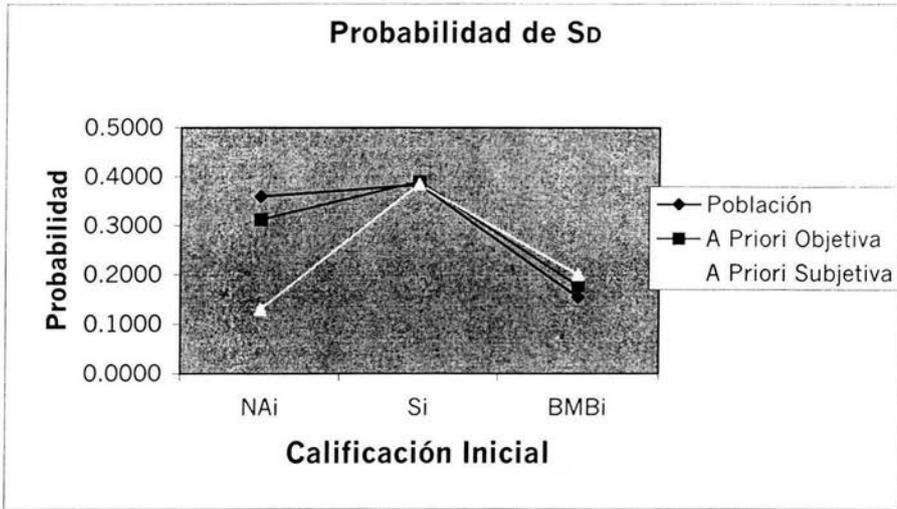


Figura 3. Probabilidad de obtener  $S_0$  dada la calificación inicial.

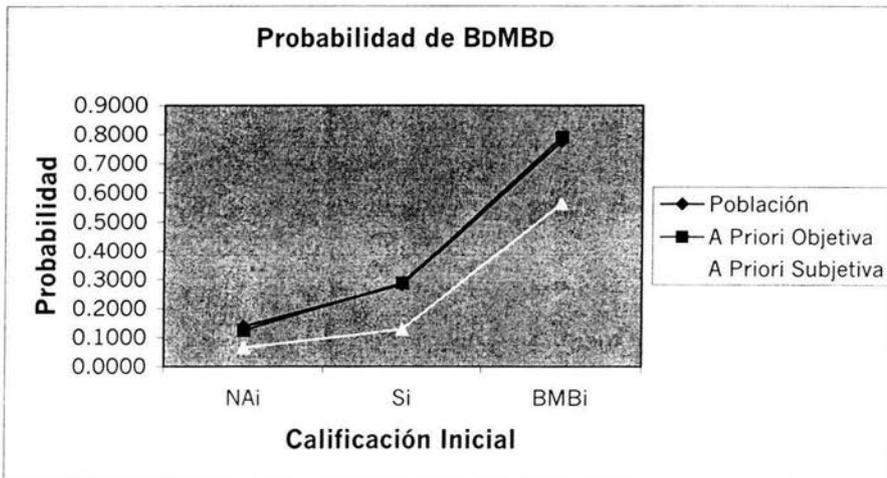


Figura 4. Probabilidad de obtener  $B_0MB_0$  dada la calificación inicial.

La figura 2 muestra que los valores estimados para obtener  $NA_D$  dada la calificación inicial del procedimiento bayesiano con una *a priori* objetiva coinciden casi totalmente con los datos reales de la población. El procedimiento bayesiano para la *a priori* subjetiva obtiene valores más elevados.

En la figura 3 es posible observar que los estimadores bayesianos objetivo y subjetivo predicen con exactitud los datos de la población para  $S_I$  y  $B_I$  o  $MB_I$ , no así para  $NA_I$ .

La figura 4 muestra una coincidencia casi total entre el estimador bayesiano basado en una *a priori* objetiva y la población.

### **Comparación gráfica de valores esperados**

En las figuras 5, 6 y 7 se muestran los valores esperados para cada calificación inicial junto con la expresión de los datos de la muestra y de la población.



Figura 5. Valor esperado para NA inicial. Como puede observarse, la predicción más exacta es la que corresponde a la a priori subjetiva y los procedimientos clásicos y objetivos ofrecen aproximadamente la misma predicción.



Figura 6. Valor esperado para S inicial. La predicción de cada uno de los procedimientos coincide.



Figura 7. Valor esperado para calificación inicial B o MB. El valor esperado para el procedimiento objetivo es el más acertado.

### Cálculo del coeficiente de correlación

Con el fin de establecer la relación entre estimadores de probabilidad y los datos de la población se eligió el coeficiente de correlación Producto-momento de Pearson. Se realizó el cálculo de acuerdo con la siguiente expresión:

$$r_{\text{cov}(x,y)} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{[\sum_{i=1}^n (X_i - \bar{X})^2][\sum_{i=1}^n (Y_i - \bar{Y})^2]}}$$

La hipótesis nula establece que no existe relación entre los estimados bayesianos de probabilidad *a posteriori* con una *a priori*

La hipótesis nula establece que no existe relación entre los estimados bayesianos de probabilidad *a posteriori* con una *a priori* objetiva y con una *a priori* subjetiva y los datos reales de la población, respectivamente. La estimación de probabilidad con una *a priori* no informativa no fue comparada debido a que presenta un mismo valor para todos los casos, lo que imposibilita una asignación de rangos adecuada. El nivel de significación elegido fue de una  $\alpha$  de 0.05 con una  $N$  de 9 y grados de libertad  $n-2$ ; el único valor significativo, de acuerdo con el nivel elegido, fue el de la comparación: población vs probabilidad *a priori* objetiva, que fue de 0.9825 (significativo a un nivel mayor de 0.001%).

## Capítulo IV

### Discusión

#### ¿Cuál es el mejor procedimiento?

El procedimiento de inferencia basado en una *a priori* no informativa presenta los valores más altos de  $\chi^2$  (ji cuadrada), lo que evidencia un menor grado de ajuste con los datos de la población; sin embargo en el cálculo del valor esperado para B o MB iniciales concuerda con la predicción objetiva y resulta de utilidad.

El procedimiento de inferencia basado en una *a priori* subjetiva es exacto para predecir las calificaciones finales más probables en tres de nueve casos, aproximado en cinco casos y totalmente erróneo en un caso. La subjetividad resulta bastante útil también. Los valores de  $\chi^2$  (ji cuadrada), en comparación con el procedimiento bayesiano para la *a priori* objetiva, aparecen más altos, lo que evidencia que el ajuste de curva es menor al que proporciona el procedimiento para la *a priori* objetiva que obtuvo resultados casi perfectos. Al procedimiento para la

*a priori* no informativa le corresponde el valor más alto de  $\chi^2$  (ji cuadrada), lo que evidencia el menor grado de ajuste con los datos.

Con referencia al estimador de valor esperado, los procedimientos objetivo y subjetivo arrojan resultados igualmente válidos, ya que predicen con exactitud en dos de tres casos, y el procedimiento no informativo exclusivamente en uno.

El procedimiento de inferencia basado en una *a priori* objetiva fue exacto para predecir las calificaciones finales más probables en ocho de nueve casos y bastante aproximado en uno. Definitivamente el procedimiento basado en la captura de datos confiables es el que predice con mayor exactitud. Los valores de  $\chi^2$  (ji cuadrada) son cercanos a cero, lo que muestra que el ajuste de curva es casi perfecto.

La correlación de 0.98 para el procedimiento bayesiano basado en la *a priori* objetiva muestra una relación muy alta, Ahora bien, la de 0.62 para el procedimiento bayesiano para la *a priori* subjetiva resulta significativo al nivel del 0.10%, lo cual es aceptable dado el nivel de incertidumbre de la información inicial.

Aunque el procedimiento bayesiano objetivo se ajusta mejor a la curva de la población, el subjetivo se muestra aceptable cuando no se

tiene tiempo de recabar información objetiva; así, el mejor procedimiento será aquél que se elija de acuerdo con la situación, el tiempo, los recursos disponibles y la función de pérdida.

La elección del mejor procedimiento, ya sea objetivo o subjetivo, depende de diversas circunstancias: 1) ¿Se dispone de información confiable? 2) ¿Se tiene conocimiento sobre cómo se comportó en el pasado el fenómeno a estudiar? 3) ¿Se cuenta con registros empíricos? Si la respuesta es afirmativa a una o más de las preguntas anteriores, la mejor opción a utilizar será una *a priori* objetiva.

Ahora bien, si 1) no se cuenta con datos y no se tienen tiempo y recursos para recabarlos. 2) el tomador de decisiones es experto o 3) el tomador de decisiones experto cree que sabe que el fenómeno se comporta como una distribución de probabilidades conocida. La respuesta es que la mejor opción es la utilización de una *a priori* subjetiva.

## Posibles usos de los procedimientos bayesianos propuestos

Si se desea implementar algún tipo de educación remedial con base en el conocimiento anticipado de las calificaciones finales más probables, los procedimientos bayesianos de probabilidad *a priori* subjetiva y objetiva constituyen opciones a elegir. Si se desea mayor seguridad en la decisión, esto es, si la función de pérdida es grande, tal vez valdría la pena recabar datos y con base en estos aplicar el método propuesto y tomar una decisión; por otro lado, si las consecuencias de “nuestra” decisión no son graves en función del dinero que se puede perder, los alumnos que puedan desertar etc., y no se cuenta con recursos o tiempo, es posible utilizar el método de estimación de probabilidad bayesiana con una *a priori* subjetiva.

En evaluación diagnóstica, formativa y sumativa, para realimentar el proceso educativo y mejorar las decisiones de administración educativa y mejorar la didáctica.

Para la estimación de la validez predictiva de diversas pruebas psicométricas. Por ejemplo, un *test* de comprensión de lectura

considerado como probabilidad *a priori* y el rendimiento en ciencias expresado en calificaciones de la asignatura en particular.

### **Limitaciones del estudio**

El cálculo de probabilidades *a priori* subjetivas es un asunto muy delicado. Existe evidencia empírica de que tomadores de decisiones inexpertos provocan resultados catastróficos que pueden involucrar funciones de pérdida grandes para los estudiantes y en general, para el sistema educativo.

El procedimiento para asignar probabilidades *a priori* subjetivas, precisamente por su carácter de subjetivo, resulta difícil de replicar. El procedimiento subjetivo que se presenta es fácilmente replicable y predice aceptablemente.

El cálculo de probabilidades *a priori* no informativas utilizando el estimador de valor esperado constituye un elemento poco válido ya que sólo logra predecir acertadamente en un caso.

Por otro lado, aunque la exactitud de la predicción del procedimiento bayesiano para una *a priori* objetiva resulte

sorprendente, en evaluación referida a un criterio es un sistema confiable debido a la consistencia en la forma de medir y de evaluar, esto es, de calificar. Si un profesor midiera el aprendizaje de sus alumnos de una forma para el primer examen y de otra para los subsiguientes, tal vez la exactitud de la predicción no fuera tan precisa.

¿Qué función describe el fenómeno? La aplicación de un procedimiento de regresión lineal no es aplicable debido a que a todas luces el fenómeno no se comporta gráficamente de esta forma. El producto final fueron tres métodos de inferencia que no constituyen una función definida entre sí.

## **Sobre la objetividad y la subjetividad**

El desarrollo de procedimientos para medir objetivamente el rendimiento de los alumnos ha sido sin duda un gran avance en el desarrollo de la psicología como ciencia experimental. Sin embargo, la posición extrema que únicamente admite datos objetivos desvía la atención hacia el procedimiento, se circunscribe en un paradigma específico de la ciencia y en particular de la psicología, y se olvida de los fines de la educación. Ya no se considera si los estudiantes resuelven problemas utilizando procedimientos novedosos sino que se limita la

medición a si obtuvo la respuesta correcta y acertó en la elección en una prueba de opción múltiple. La elicitación de procesos intelectuales complejos es poco frecuente y la expresión escrita de los estudiantes no es tomada en cuenta, en la mayor parte de los casos.

Thomas Bayes afirma que la probabilidad es la creencia personal sobre la ocurrencia de un evento. Platón considera que la experiencia personal no es sino una sombra desdibujada de la realidad. Descartes piensa que los sentidos nos engañan con frecuencia y que no se puede confiar en el pensamiento ya que frecuentemente nos conduce al error. Bacon, los empiristas y los antiguos conductistas desecharían totalmente la utilización de creencias subjetivas pues su concepto de realidad, siguiendo a Platón, se limita a la medición objetiva de un conjunto de sombras distorsionadas de la realidad.

Por otra parte, Hume y Einstein se dan cuenta que el problema no es tan sencillo y que de las experiencias sensibles se deriva la conceptualización y la creencia subjetiva de que ésta es la realidad, desarrollándose una interacción que permea e influencia ambos campos: el de la realidad objetiva y el de la subjetiva. Las creencias subjetivas y las probabilidades no son otra cosa que una expresión subjetiva-objetiva de nuestra ignorancia.

El escepticismo de Hume es superado por Kant que establece un punto de apoyo para el desarrollo del conocimiento científico. Si bien no es posible acceder a la realidad como tal, sí podemos saber de las cosas en cuanto como se nos presentan, es decir, en cuanto fenómenos. La realidad objetiva no es absoluta, así como la subjetiva tampoco. La una es necesaria para la otra y asimismo la contiene. Las creencias subjetivas están basadas en la experiencia y en el procesamiento de la información que desarrolla el ser humano. Cuando Einstein postuló la Teoría de la Relatividad, hizo uso de su intuición, los experimentos que la corroboraron empírica y objetivamente pudieron realizarse años más tarde de enunciada la teoría. Sin embargo, Einstein sabía de todos los conocimientos fundamentales de la física de su tiempo que, junto con los problemas que la física no había podido resolver, hicieron que sus creencias subjetivas enunciaran una de las teorías científicas más importantes de nuestro tiempo.

Definitivamente, la utilización de las creencias subjetivas bayesianas es legítima y conveniente. La verificación experimental de las mismas con datos objetivos también.

## Líneas de investigación posibles

Actualmente la estadística bayesiana se utiliza en numerosos campos debido a la exactitud de sus predicciones. En psicología ya existe una línea abierta de investigación para evaluación formativa en secuencias de enseñanza programada individualizada, pero también podría utilizarse para el diagnóstico en psicología clínica y medicina, por ejemplo para decidir si es conveniente que el paciente, dado su estado de depresión, tome fármacos o no o para decidir el tipo de terapia que más le convenga.

Otros usos serían la toma de decisiones en programas de entrenamiento para decidir si es conveniente que el estudiante pase al siguiente nivel o no, la selección de estudiantes que resuelven un examen de admisión, la evaluación de dominio de una unidad en educación a distancia y desarrollo de procedimientos para medir estrategias intuitivas en la solución de problemas, entre otros.

También es posible diseñar investigaciones en diferentes campos que inicien sin información alguna, como en el caso de una *a priori* no informativa y utilicen la *a posteriori* resultante como la *a priori* del segundo paso del proceso, incluir datos objetivos o subjetivos y

continuar las mediciones y los cálculos de probabilidad incrementando la exactitud de la predicción en cada etapa del proceso.

## Conclusiones

El presente estudio reafirma las conclusiones preliminares del estudio de Rosas Mercado en el sentido de que la probabilidad bayesiana con base en una *a priori* objetiva es un método válido de predicción de las calificaciones finales a partir de una calificación y lo perfila como el más exacto en la predicción ante diferentes tipos de incertidumbre inicial. También establece que los procedimientos subjetivos predicen con cierto grado de exactitud. No obstante, pone en duda si las posiciones divergentes de Savage y Jeffreys con relación a la efectividad de las concepciones subjetiva y objetiva son adecuadas, pero, ¿cuál es la mejor? La aproximación subjetiva al problema tal vez no resulta tan acertada con un tomador de decisiones inexperto o que sepa muy poco del fenómeno del que se ocupa; sin embargo, una vez que el procedimiento "subjetivo" se estandarice para un problema particular, los riesgos disminuyen.

El marco de inferencia bayesiano utilizado como evaluación diagnóstica, formativa o para asignar calificaciones, integra datos

objetivos y proporciona predicciones válidas, involucra valoraciones subjetivas de expertos y también proporciona inferencias certeras.

La utilización de creencias subjetivas constituye un procedimiento válido que predice acertadamente y la utilización del teorema de Bayes con base en registros de calificaciones anteriores resulta un método de predicción más exacto.

## Nomenclatura

### Griegas:

- $\alpha$  : nivel de confianza deseado.
- $\sigma^2$  : varianza de la población.
- $\mu$  : media de la población.
- $\chi^2$  : ji o chi cuadrada.
- $\Sigma$  : operador sumatoria.

### Mayúsculas:

- (A | B)** : condicionalidad; evento A, dado B.
- C** : número de clases.
- C** : calificación.
- F** : frecuencia esperada.
- G.L.** : grados de libertad.
- N** : número total de elementos en la población.
- P** : probabilidad.
- S** : espacio muestral.

### Minúsculas:

- d** : desviación permitida como aceptable en la predicción de la calificación.
- f** : frecuencia observada.
- n** : número total de elementos en la muestra.
- p** : proporción de casos aprobados en la premuestra.
- r** : número de muestras.
- t** : valor de desviación estándar correspondiente al nivel ( $\alpha$ ) de confianza deseado.

### Subíndice:

- D** : definitiva.
- I** : inicial.
- i** : iésimo valor.
- k** : kaésimo valor; valor específico de un índice variable.
- M** : marginal.

## GLOSARIO

*Aleatorios, resultados*: Resultados inciertos de un experimento.

*Colectivamente exhaustivo, resultado*: Cuando la unión de todos los resultados es igual al espacio muestral.

*Confiabilidad*: Consistencia de la herramienta de medición. Precisión con que mide la prueba. Consistencia de los datos en mediciones repetidas (dadas las mismas condiciones) con el mismo instrumento.

*Curva normal de probabilidad*: Gráfica ideal de distribuciones de probabilidad cuya validez crece cuando N es muy grande y que sigue la expresión:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(\frac{-x^2}{\sigma^2}\right)^2}$$

*Desviación estándar*: Medida de dispersión o variabilidad de las puntuaciones en un grupo determinado. La desviación estándar es la raíz cuadrada del promedio de las desviaciones cuadradas respecto de la media aritmética del grupo. Se expresa mediante la relación siguiente:

$$\sigma = \sqrt{\sum_i (x_i - u)^2 p_i}$$

*Diseño de la prueba:* Parte que trata los pasos destinados a estimar el grado de probabilidad en que las observaciones concuerden con las hipótesis presentadas.

*Distribución:* Arreglo relativo de un conjunto de números; un conjunto de valores de una variable y las frecuencias de cada valor.

*Distribución de frecuencias:* Ordenamiento de un conjunto de puntuaciones correspondientes a un grupo de individuos en el que los valores de las puntuaciones se disponen en orden descendente y se indica el número de individuos que obtiene cada puntuación.

*Ente:* Aquello que existe.

*Error de medición:* Diferencia que existe entre la puntuación hipotética y la puntuación “verdadera” de un individuo.

*Espacio muestral:* Conjunto universal formado por todos los elementos que representan todos los posibles resultados de un experimento. Se representa con la letra S.

*Estadística inferencial:* Es un cuerpo estructurado de técnicas que tratan de inferir las propiedades de un conjunto muy grande de datos a través de la inspección de una muestra del conjunto, obteniendo inferencias de propiedades particulares de las muestras y aplicándolas a propiedades particulares de las poblaciones.

*Estocásticos, resultados:* Resultados dudosos de un proceso de cambio.

*Evaluación:* Proceso que incluye la interpretación del significado de los puntajes producto de las mediciones, y las resoluciones para precisar si son los correctos o no para determinados objetivos.

*Evaluación diagnóstica:* Aquélla que recaba información que trata el carácter de alguna dificultad o carencia en un individuo, factor, esquema, desarrollo, etc.

*Evaluación formativa:* Aquélla que tiene como finalidad proporcionar la información que orienta al perfeccionamiento del ente que se está evaluando.

*Evaluación sumativa:* Aquélla que tiene como intención obtener una determinación evidente que puede ser una calificación binaria o múltiple en relación a un esquema, individuo o factor con base en su importancia, generalmente en relación con algún punto fundamental.

*Experimento:* Prueba física que se efectúa x número de veces bajo condiciones similares para conseguir un conjunto de resultados.

*Feedback:* (Realimentación) Función que tiene como finalidad controlar el rendimiento del sistema mediante la comparación del resultado logrado con los criterios de rendimiento establecidos, facilitando formaciones esenciales que admitan ajustes progresivos para suprimir las diferencias que pudieran existir entre los criterios y el producto.

*Fenómeno:* Lo que puede percibirse de la realidad a través de los sentidos, lo que se presenta ante ellos. Cualquier manifestación de

actividad que se presenta en la naturaleza y que es percibida por los sentidos o instrumentos de medición.

*Función:* Relación en que a cada elemento de un conjunto (el rango) se le hace corresponder con un elemento de otro conjunto (el dominio). La función entre conjuntos se denota con  $F$  ó  $f$ .

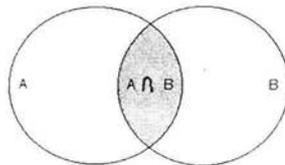
*Grados de libertad:* Número de categorías o clases cuyos valores pueden sufrir variaciones arbitrarias.

Para una muestra: Grados de libertad= $C-1$ ; donde  $C$ =número de categorías o clases.

Para  $r$  muestras: Grados de libertad= $(r-1)(C-1)$

*Incertidumbre:* Grado de conocimiento sobre la ocurrencia de un evento. La incertidumbre tendrá un valor de cero cuando el resultado de un evento sea seguro, en este caso, la probabilidad asociada será 1.

*Intersección:* Conjunto formado por los elementos comunes a ambos conjuntos,  $A$  y  $B$ . Su símbolo es  $\cap$ .



*Iterativos, procedimientos:* Proceso repetitivo que considera los resultados obtenidos a lo largo del mismo para aproximarse sucesivamente a la solución del problema.

*Intuición:* Conocimiento inmediato de un evento, idea o hecho sin el concurso aparente del razonamiento.

*Halo o aura, error de:* Preferencia de algunos evaluadores a apoyar sus estimaciones con base en una opinión general favorable de un individuo, sin tomar en cuenta rasgos específicos.

*Hecho o evento:* Cualquier subconjunto del espacio muestral.

- Simple: Evento formado por un punto en el espacio.
- Compuesto: Evento compuesto por dos o más puntos.

*Hechos disjuntos:* Eventos que no tienen eventos en común. Así su intersección es  $A \cap B = 0$ .

*Hipótesis:* Respuesta tentativa a un problema o pregunta. Afirmación que relaciona dos o más variables.

*Hipótesis alternativa:* Aquélla que establece el efecto de la variable independiente sobre la variable dependiente.

*Hipótesis nula.* Aquélla que establece un efecto nulo de la variable independiente sobre la variable dependiente, por lo que no se esperan diferencias en los valores de la variable dependiente bajo diferentes condiciones experimentales.

*Medición:* Proceso que consta de objetivos, construcción de instrumentos de medición, administración y asignación de puntos a objetos sujetos a pruebas.

*Modelo:* Representación de circunstancias reales mediante variables.

*Objetivo:* “Es un propósito expresado en un enunciado que describe el cambio propuesto en el alumno cuando haya terminado exitosamente una experiencia de aprendizaje” (Mager, 1969).

*Permutación:* Arreglo de un conjunto de  $n$  objetos en un orden dado. El número de permutaciones de  $n$  objetos tomados  $r$  a la vez es:

$${}^n P_r = \frac{n!}{(n-r)!}$$

*Probabilidad:* Medida que expresa la incertidumbre y sus riesgos.

- *Objetiva:* Creencia de un individuo basada en observaciones y/o mediciones de los fenómenos. Creencia de un individuo compartida por otras personas y contrastada con observaciones y/o mediciones.
- *Subjetiva:* Creencia de un individuo sobre la ocurrencia de un evento o conjunto de ellos basada en su experiencia, el el razonamiento o en la intuición.
- *Bayesiana:* Grado de confianza en la posibilidad de que un hecho ocurra.
- *A priori:* Creencia de un individuo a cerca de la ocurrencia de un evento, antes de efectuarse las observaciones.
- *A posteriori:* Creencia de un individuo acerca de la ocurrencia de un evento, después de efectuadas las observaciones.
- *Condiciona:* Aquélla que indica que un evento B sucede si y sólo si un evento A ya ocurrió. Se representa como:  $P(B|A)$ .
- *Marginal:* Suma de probabilidades de las intersecciones de cada uno de los valores de la variable independiente con los distintos valores

de la variable dependiente. Considerando que X es la variable independiente y que Y es la variable independiente, su fórmula general sería:

$$P(X_k) = \sum_{i=1}^n P(Y_i \cap X_k)$$

*Punto muestral:* Cada uno de los resultados de un experimento.

*Puntuación bruta:* Puntuación que se expresa con las unidades obtenidas durante la aplicación de la prueba.

*Razonamiento:* Acción de proporcionar las bases en que se fundamenta un evento.

*Razón insuficiente, principio de Bernoulli:* Señala que los resultados son igualmente probables en un experimento al azar.

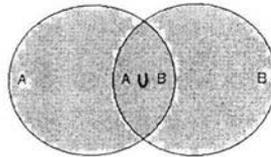
*Resultado aleatorio o estocástico:* Resultados inciertos de un experimento.

*Resultados inciertos:* Cuando en un proceso de cambio puede haber más de un resultado posible.

*Sistema:* Ente o conjunto de entes compuesto de partes organizadas en un todo destinado a lograr uno o más propósitos generales o específicos.

*Teorema:* Toda proposición que para ser aceptada como verdad, requiere ser demostrada.

*Unión:* Es el evento que contiene todos los elementos que pertenecen a A o B o ambos. Su símbolo es  $\cup$ .



*Validez:* Eficiencia de una prueba para representar, explicar o predecir el atributo que se mide a través de ella.

*Valor esperado o expectativa matemática:* “El valor esperado de una variable aleatoria X es un promedio ponderado de los valores que puede asumir X con probabilidades para los valores de X como pesos” (Chou, 1977).

*Variable:* Símbolo que representa un miembro no específico de algún conjunto.

*Variable aleatoria:* Función numérica real definida en un espacio muestral.

*Variable aleatoria continua:* Aquella cuyo rango es un conjunto de números reales y que su espacio muestral contiene un número finito o infinito de puntos que forman una continuidad.

*Variable aleatoria discreta:* Aquella cuyo rango es un conjunto de números reales y que su espacio muestral es discreto.

### *Referencias Bibliográficas*

- Artis, M. J. & Zhang, W. (1990, octubre). Bvar forecasts for G - 7. *International Journal of Forecasting*, 6, 349 -362.
- Bain, L. J. & Engelhardt, M. (1991, octubre). Reliability test plans for one-shot devices based on repeated samples. *Journal of Quality Technology*, 23, 304-311.
- Bacon, F. (1952, Reimpreso) Novum Organon, Aphorisms Concerning the Interpretation of Nature and the Kingdom of Man, en The Great Books, Volumen 30, Encyclopaedia Britannica, Inc.: The University of Chicago, Estados Unidos de América, 107-195.
- Bayes, T. (1763/1958, diciembre). Essay towards solving a problem in the doctrine of chances. (reimpreso) *Biométrica*, (Vol. 45), X, 299, 306.
- Biblia (La). Jueces, 12:5-6, Versión del Rey Juan.
- Bloom, B. S., Thomas, H. J. & Madaus, G. F. (1975). *Evaluación del aprendizaje*, (Vol I), Buenos Aires, Argentina: Troquel, 135-143.
- Box, G. E. & Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Menlo Park, E.U.A.: Addison -Wesley Publishing Company, 41-58.

- Castro, L. (1976). *Diseño experimental sin estadística*. México, D.F.: Trillas.
- Chou, Y. L. (1977). *Análisis estadístico*. México, D.F.: Interamericana. 118-124.
- Clements, M. P., Mizon, G., Kirchgassner, G. & Sims, C.A. (1991, mayo). Empirical analysis of macroeconomic time series: Var and structural models; comments". *European Economic Review*, 35, 887 – 932.
- Cochran, W. G. (1971). *Técnicas de muestreo*. México, D.F.: CECSA, 109-110.
- Cochran, W. G. & Cox, G. M. (1981). *Diseños experimentales*. México, D.F.: Trillas, 109-110.
- Cohen, R.J. & M.E.E. Swerdlick (2000). *Pruebas y Evaluación Psicológicas*, México, D.F.: Mc. Graw Hill, 48-49, 137-139, 139,139.
- Collinson, P. O., Cramp, D. G., John, Ms. K., Canepa-Anson, R., O'Donell, J., & Carson, E. R. (s.f.). *Bayesian evaluation*. Servidor:city.ac.uk  
Directorio:/mim/Archivo:technology.
- Crémilleux, B., Gao, M., Madelaine, J. & Zreik, K. (s.f.) (2001, mayo) Discovering browsing paths on the web. Servidor:europia.org  
Directorio:/communication/CAPS/Archivo:Abstract.
- Curda L. K. & Curda, S. (2001, 11 de septiembre). Formative evaluation proceses in two web-based courses: making just-in-time improvements [en red]. Servidor:aect.org/Directorio:Events/Atlanta/Presentations /archivo:detail.asp?ProposalID=516

Descartes, R. (Reimp 1952) Rules for the direction of mind, en The Great Books, Volumen 31, Encyclopaedia Britannica, Inc.: The University of Chicago, Estados Unidos de América , 3.

Díaz, B. A. (Compilador). (2000). *El examen, textos para su historia y debate*. México, D.F.: Universidad Nacional Autónoma de México, 12-14, 12-13, 20.

*Diccionario Larousse Ilustrado*, (1999). México, D.F.: Editorial Larousse.

*Diccionario Enciclopédico* (2 000). México, D.F.: Olimpia.

Einstein, A. (1969/Reimp. 1982). La relatividad. México, D.F., Grijalbo. 185.

Forester, L. (1973). *La estadística en la toma de decisiones*. Bilbao, España: Deusto. 67-81.

Freeman, H. (1970.) *Introducción a la inferencia estadística*. México, D.F.: Trillas.

Glass, G. V. (1977, diciembre). *Standards and Criteria*. Disponible en pdf. Servidor:wmich.eduDirectorio:/evalctr/pubs/ops/archivo:ops10

Glass, G. V. & Stanley J. C.(1970). *Statistical Methods in Education and Psychology*. Englewood Cliffs, New Jersey, E.U.A.: Prentice Hall, Inc, 7, 8-14.

- Gimeno S. J. (1999). La evaluación en la enseñanza. En Sánchez, P.F., Compilador (1999). *Selección de lecturas. Asignatura: Evaluación educativa*, (p 7). Maestría en Ciencias de la Educación Superior. Área de Estudios sobre Educación Superior, Vice-Rectoría Docente, Matanzas, Cuba: Universidad de Matanzas.
- González Rey, F. L.(2000). *Investigación cualitativa en psicología: Rumbos y desafíos*, México, D.F.: International Thompson Editores. 18-22.
- Gregory, R. J. (2001). *Evaluación Psicológica*. México, D.F.: Mc. Graw Hill, 34, 20, 34, 9, 14-17.
- Hambleton, R. K., Swaminathan H., Algina, J. (1975) Some Contributions to the Theory and Practice of Criterion Referenced Testing. En *Advances in Psychological and Educational Measurement* (pp 51-62) Londres: Ed. N. De Gruijter.
- Haskell, R. E. (2000) A reply to Haskell and to Skate. Disponible en [html.Servidor:epaa.asu.edu/directorio/epaa\\_archivo:v5n8c3](http://html.Servidor:epaa.asu.edu/directorio/epaa_archivo:v5n8c3)
- Hoeting, J. A., Madigan, D., Raftery, A. E. & Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. Disponible en [pdf Servidor: stat.washington.edu/Directorio:/www/research/online/archivo:hoeting1999](http://pdf.Servidor:stat.washington.edu/Directorio:/www/research/online/archivo:hoeting1999)
- Houston, W. M. & Woodruff, D. J. (1997, agosto) Empirical bayes estimates of parameters from the logistic regression model, *Act Research Report Series*, 97 - 6.

- Hume, D. (1952, Reimpreso). Concerning human understanding., The Great Books, Volumen 36, Encyclopaedia Britannica, Inc.: The University of Chicago, Estados Unidos de América, 447-466, 469.
- Jackson, P. H. (1975). The philosophy and methodology of bayesian inference, en *Advances in Psychological and Educational Measurement* (3-16) Londres: Ed. N. De Gruijter.
- Jefreys, S. M. (1961). *Theory of probability* (Tercera edición). Oxford, Inglaterra: Clarendon Press, en Box, G. E. & G. C. Tiao (1973). *Bayesian Inference in Statistical Analysis*. E.U.A.: Addison -Wesley Publishing Company, pp 41-58.
- Judges, V.A. (2000). La evolución de los exámenes en Díaz, B.A. (Compilador). (2000). *El examen, textos para su historia y debate*. México, D.F.: Universidad Nacional Autónoma de México, (pp 34, 40-41, 33 y34, 35 y 37, 45).
- Kant, E. (1952, Reimp). *Crítica de la razón pura*. México, D.F.: Número 23. Colección "Sepan Cuantos...", Porrúa. 355.
- Koedinger, K. R. (2000). Intelligent tutoring goes to school in the big city. Disponible en pdf Servidor:cs.ubc.ca/Directorio:~conati/532b2000/papers/Archivo:algebratutor/
- Koedinger, K. R., Hadley, W. H. & Mark, M. A. (2000). Research statement. Disponible en html. Servidor:act.psy.cmu.edu/Directorio:ACT/ people/ Archivo: koedingerResearch

- Kumar, V., McCalla, G. & Greer, J. (2000). Graduate student's projects. Disponible en pdf. Servidor:cs.usask.ca/Directorio:research\_groups/aries/Archivo:gradprojects1
- Landsheere, D. G. (1973). *Evaluación continua y exámenes*. Buenos Aires, Argentina : El Ateneo, VII-IX.
- Lewis, C. (1975). Marginal distributions for the estimation of proportions in m groups. *Psychometrika*, 28.
- Mager, R. F. (1965). *La confección de los objetivos de enseñanza*. La Habana, Cuba : Año de los Diez Millones.
- Magnusson, D. (1966). *Teoría de los tests*. México, D.F.: Trillas. 26-31.
- Maduako, K., Cramp, G. y Carson, E.R. (2001) *Measurement and Information in Medicine*. Londres, Servidor:city.ac.uk/Directorio:mim/Archivo/technology.htm
- Mariás, J. (1941). *Historia de la filosofía*. Madrid, España: Revista de Occidente, S.A., 204-216, 307-319.
- Mislevy, R. J. Steinberg, L. S. & Almond, R. G. (1999, enero). On the roles of task model variables in assessment design. *CSE Technical Report 500. National Center for Research on Evaluation, Standards, and Student Testing (CRESST)*.
- Mood, A. M. & Graybill F. A. *Introducción a la teoría de la estadística*. Madrid, España: Aguilar.

- Morgan, B. W. (1971). *Introducción a los procesos bayesianos de decisión estadística*, Madrid, España: Paraninfo.
- Novick, M. R., Jackson, P. H., Thayer, D. T. & Cole, N. S. (1971, abril) Applications of bayesian methods to the prediction of educational performance. *Act Research Report*, 42.
- Novick, M. R. (1970, noviembre). Bayesian considerations in educational information systems. *Act Research Report*, 38.
- Novick, M. R. (1970). *Bayesian Considerations in Educational Informations Systems*. Nueva York, E.U.A. : Educational Testing Services, 77-89.
- Platón, (s. IV a.C./1952). The dialogues of Plato. The Great Books, Volumen 7, Encyclopaedia Britannica, Inc.: The University of Chicago, E.U.A., 388-399.
- Phillips, L. D. (1974). *Bayesian Statistics for Social Scientists*. Nueva York, E.U.A.: Thomas and Crowell Company.
- Rouanet, H., Lépine, D. & Pelnard-Considère, J. (1975). Bayes-fiducial procedures as practical substitutes for misplaced significance testing: an application to educational data, *Advances in Psychological and Educational Measurement*. (pp 33-50). Londres: Ed. N. De Gruijter.
- Rosas Mercado, A. (1984). Estimación bayesiana de calificaciones. XXVI *Congreso Internacional de Psicología*. Acapulco, Guerrero, México.

- Sánchez P. F. Compilador. (1999). *Selección de lecturas en evaluación educativa*. Matanzas, Cuba: Maestría en Ciencias de la Educación Superior. Área de Estudios sobre Educación Superior, Vice-Rectoría Docente, Universidad de Matanzas.
- Savage, L. J. (1954). *The foundations of statistics*. (pp 17-32) New York: Wiley, (pp 17-32) citado por Box, G. E. & G. C. Tiao (1973). *Bayesian inference in statistical analysis*. E.U.A.: Addison -Wesley Publishing Company, 14.
- Siegel, S. (1978). *Estadística no paramétrica*. México, D.F.: Trillas.
- Stanley, J.C. & Hopkins, K.D. (1972). *Educational and Psychological measurement and evaluation*. Cuarta edición. New Jersey, USA: Prentice-Hall, 4 y 5.
- Starch, D. (1913). Reliability and distribution of grades. *EUA: Science*, 38. Citado por Stanley, J.C. y Hopkins, K.D. (1972). *Educational and Psychological Measurement and Evaluation*. Cuarta edición. USA: Prentice-Hall.
- Syvantek, D. J., O'Connell, M. S. & Baumgardner, T. L. (1992, junio). Applications of bayesian methods to evaluation and decision making. *Human Relations*, 45, 621-636.
- Theall, M. (1997) *On drawing reasonable conclusions about students ratings of instruction: a reply to Haskell and to Stake*. Servidor: epaa.asu.edu/Directorio: epaa/Archivo: v5n8c2

- Van Den Broeck, Koop, G., Osiewalski, J. & Steel, M. F. (1994, abril). Stochastic frontier models: a bayesian perspective. *Journal of Econometrics*, 61, 273 - 303.
- Vanlehn, K. & Zhendong, N. (2001). Bayesian student modeling, user interfaces and feedback: a sensitivity analysis. Disponible en html. Servidor:cbl.leeds.ac.uk/Directorio:ijaied/abstracts/Vol\_12Archivo:vanlehn
- Vanlehn, K. & Martin, J. (1997). Evaluation of an assessment system based on bayesian student modeling. Disponible en html. Servidor:cbl.leeds.ac.uk/Directorio:ijaied/abstracts/Vol8/vanlehn
- Vos, H. J. (2000). A bayesian procedure in the context of sequential mastery testing. Disponible en pdf. Servidor:uv.es/Directorio:/psicologica/archivo:articulosly2.00/vos
- Willemsen, E. (1974). *Understanding statistical reasoning how to evaluate research literaure in the behavioral sciences*. E.U.A.: Freeman and Co.
- Wilmick, F. W. & Nievels, K. A. (1982). A bayesian approach for estimating the knowledge of item pool in multiple choise tests". *British Journal of Mathematical and Statistical Psychology*, 35 (1), 90-101.
- Yoe, C. E. (1996, septiembre) Incorporating risk and uncertainty into environmental evaluation: an annotated bibliography. Disponible en pdf.Servidor:wrsc.usace.army.mil Directorio:iwr/pdf/archivo:96r09

Zapata, D. & Greer J., (2002) Visualizing Learner Models, Graduate students project. Disponible en html. Servidor:cs.usask.ca/Directorio:research/research\_groups/aries/archivo:gradprojects1

Ziomek, R. L. & Andrews, K. M. (1996, noviembre). Predicting the college grade point averages of special-tested students from their act assessment scores and high school grades. *Act Research Report Series, 96-97.*