LIBRO DE ACTAS

CACIC 2017

XXIII CONGRESO ARGENTINO DE CIENCIAS DE LA COMPUTACIÓN

XXIII Congreso Argentino de Ciencias de la Computación (CACIC 2017)

La Plata (Buenos Aires), 9 al 13 de octubre 2017

Organizadores

Red de Universidades con Carreras en Informática – RedUNCI.

Facultad de Informática - Universidad Nacional de La Plata

XXIII Congreso Argentino de Ciencias de la Computación CACIC 2017: libro de Actas.
Editado por Armando De Giusti; Patricia Pesado. - 1a ed.
La Plata: Universidad Nacional de La Plata. Facultad de Informática, 2017.
Libro digital, PDF
Archivo Digital: descarga y online
ISBN 978-950-34-1539-9

ISBN 978-950-34-1539-9
9 789503 415399

1. Computación. 2. Actas de Congresos. I. De Giusti, Armando, ed. II. Pesado, Patricia, ed.
CDD 004

Formulación de un Examen Óptimo

Enrique E. Tarifa[1,2], Sergio L. Martínez[1], Samuel Franco Domínguez[1] y Jorgelina F. Argañaraz[3]

[1]Facultad de Ingeniería, Universidad Nacional de Jujuy
[2]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
[3]Facultad de Humanidades y Ciencias Sociales, Universidad Nacional de Jujuy
4600 San Salvador de Jujuy, Argentina
{eetarifa, smartinez, sfrancodominguez}@fi.unju.edu.ar
jfarganaraz@hotmail.com

Resumen. El objetivo del presente trabajo es formular un examen óptimo para una materia dada. Para ello, primero se modela la probabilidad de que un estudiante apruebe el examen en función de la cantidad de unidades que evalúa el profesor. El modelo de simulación que se presenta en este trabajo se desarrolla realizando un análisis probabilístico. Luego, se define un examen óptimo como aquel que adjudica la nota que merece el estudiante, y que puede tomarse respetando las limitaciones de tiempo y esfuerzo que afectan al docente; en consecuencia, en un examen óptimo, aprueban los que merecen aprobar, y desaprueban los que no merecen aprobar. En base a esta definición, empleando el modelo de simulación, se formula un modelo de optimización del tipo INLP que determina la cantidad de unidades a evaluar que maximiza la probabilidad de tomar un examen óptimo.

Palabras clave: Evaluación académica, Análisis probabilístico, Optimización

1 Introducción

La evaluación es un tema crítico en cualquier institución, y particularmente en las educativas. Según Frola y Velásquez [1], el proceso de evaluación implica la obtención de información, la emisión de juicios una vez procesada la información, y la consecuente toma de decisiones tendientes a la mejora de los procesos y servicios. La evaluación puede ser cualitativa o cuantitativa, siendo la primera preferida para evaluar el aprendizaje, mientras que la segunda es elegida para medir los conocimientos que al final de un periodo el estudiante retiene.

A pesar de la importancia que tiene la evaluación, es una cuestión que no ha sido todavía adecuadamente resuelta en el ámbito educativo, con el consiguiente impacto negativo sobre la formación de los estudiantes. En efecto, Trillo Alonso y Porto Currás [2] analizaron la percepción que tenían los estudiantes sobre la evaluación en la Facultad de Ciencias de la Educación de la Universidad de Santiago de Compostela durante el curso 1997-1998, y concluyeron que, para los estudiantes, la evaluación no cumple con los objetivos que debería tener. Frente a este resultado, los autores plantean que, si esto ocurre en la Facultad de Ciencias de la Educación, no es probable un escenario mejor en otras facultades. Por otra parte, los resultados son más preocupantes si se considera que dichos estudiantes serán futuros profesionales de la educación.

Las TIC, además de influir en otros aspectos de la educación, también influyen en la forma de evaluar, abriendo nuevas posibilidades al automatizar las correcciones, calcular índices estadísticos, realizar histogramas [3]. Estas funciones favorecen una mejor evaluación; pero no brindan una solución acabada; es más, abren nuevos temas. Por ejemplo, los cuestionarios de opciones múltiples, al poder ser contestados empleando el azar, plantean el problema de corregir la nota del cuestionario considerando la posibilidad de que el estudiante haya contestado probando suerte [4].

Huapaya *et al.* [5] plantean que, para realizar una evaluación justa, además de la nota de un examen, se deben considerar otros aspectos: promedios de notas del estudiante, promedio de la clase, evolución de las notas del estudiante. Estos datos son procesados por un sistema experto con lógica difusa para diagnosticar el nivel de conocimiento del estudiante. El empleo de un sistema experto tiene la ventaja de quitar subjetividad a la evaluación, produciendo evaluaciones uniformes; pero está limitado por el conocimiento de los expertos y el proceso de adquisición del conocimiento [6].

Un aspecto importante poco investigado de las evaluaciones es el rol del profesor. El profesor, al diseñar la evaluación, toma varias decisiones: la cantidad de preguntas a realizar, los temas cubiertos por las preguntas, la nota de aprobación (si no está fijada por la institución). Las decisiones que el profesor tome en esta etapa causan un profundo impacto en el resultado de la evaluación. Por ese motivo, el presente trabajo tiene como objetivo analizar los efectos que dichas decisiones tienen sobre la evaluación. Para ello, se define un examen óptimo como aquel que adjudica la nota que merece el estudiante, y que puede tomarse respetando las limitaciones de tiempo y esfuerzo que afectan al docente; en consecuencia, en un examen óptimo, aprueban los que merecen aprobar, y desaprueban los que no merecen aprobar. En base a esta definición, se formula un modelo de optimización que determina la cantidad de unidades a evaluar que maximiza la probabilidad de tomar un examen óptimo.

2 Modelo de Simulación

2.1 Formulación del Problema

El problema inicial que se aborda en este trabajo consiste en estimar la probabilidad que tiene un estudiante que rinde un examen de una materia compuesta por *UM* unidades de evaluación —en adelante llamadas unidades de la materia— cuando el estudiante estudia *UE* unidades y el profesor toma *UT* unidades. El examen se aprueba con una nota igual o superior a *NA*, con nota máxima igual a *UM*. Las unidades de evaluación de la materia representan el grado de detalle en el que el profesor desglosa la materia para realizar la evaluación. En grado de detalle creciente, las unidades de evaluación pueden ser unidades del programa de la materia, temas de la materia, subtemas de la materia. La respuesta de un estudiante a una unidad de evaluación puede ser correcta o incorrecta, no se consideran grados intermedios. Se supone que todas las unidades de evaluación tienen el mismo grado de dificultad tanto para estudiarlas como para evaluarlas.

2.2 Solución Analítica

Dada la formulación planteada, para determinar la probabilidad de que el estudiante conteste correctamente una cantidad particular u de preguntas, se puede aplicar la distribución hipergeométrica [7], una distribución discreta relacionada con muestreos aleatorios sin reemplazo. Efectivamente, para el problema que se está analizando, se tiene una población de *UM* elementos que pertenecen a dos categorías: las unidades que el estudiante sabe y las unidades que no sabe; *UE* unidades pertenecen a la primera, y *UM-UE* unidades pertenecen a la segunda. Definidas esas categorías, la distribución hipergeométrica permite calcular la probabilidad de obtener u elementos de la primera categoría en una muestra sin reemplazo de *UT* elementos de la población original; es decir, la probabilidad $P(U = u)$ de que el estudiante conteste bien u preguntas de las *UT* realizadas por el profesor:

$$Pu(u) = \frac{\binom{UE}{u}\binom{UM-UE}{UT-u}}{\binom{UM}{UT}} \tag{1}$$

La nota que obtendrá ese estudiante será:

$$N(u) = \text{Redondeo}\left(NM\,\frac{u}{UT}\right) \tag{2}$$

Se deduce también que el valor mínimo de U es:

$$U_{\min} = \max\left(0, UT - (UM - UE)\right) \tag{3}$$

mientras que el valor máximo que puede alcanzar U es:

$$U_{\max} = \min\left(UE, UT\right) \tag{4}$$

A partir de la distribución probabilística de U, se deduce la probabilidad *PA* de que el estudiante apruebe la evaluación al obtener una nota $N(u)$ igual o superior a *NA*:

$$PA = \sum_{u=U_{\min}}^{U_{\max}} H\left(N(u) - NA\right) Pu(u) \tag{5}$$

donde $H(.)$ es la función escalón:

$$H(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases} \tag{6}$$

Por otra parte, la probabilidad de que el estudiante obtenga la nota n en el examen $P(N = n)$ se calcula de la siguiente manera:

$$PN(n) = \sum_{u=U_{\min}}^{U_{\max}} H\left(N(u) - n\right) H\left(n - N(u)\right) Pu(u) \tag{7}$$

es decir, que la probabilidad de n se obtiene sumando las probabilidades de los u con $N(u) = n$.

Finalmente, se define el valor de la nota justa NJ, como la nota que merece el estudiante que estudió UE unidades:

$$NJ = \text{Redondeo}\left(NM \frac{UE}{UM} \right) \tag{8}$$

Si esta nota es igual o superior a NA, el estudiante merece aprobar; de lo contrario, no merece aprobar.

3 Caso de Estudio

Para el análisis que sigue, se adopta como caso de estudio una materia con $NM = 10$, $NA = 4$ y $UM = 10$. En este caso, NJ será igual a UE, por lo que la cantidad mínima que puede estudiar el estudiante para merecer aprobar es 4 unidades.

4 El Examen Óptimo

Como se planteó en la introducción del presente trabajo, se define como examen óptimo a aquel que adjudica la nota que merece el estudiante, y que puede tomarse respetando las limitaciones de tiempo y esfuerzo que afectan al docente (tiempo de realización del examen, tiempo de corrección, etc.); en consecuencia, en un examen óptimo, aprueban los que merecen aprobar, y desaprueban los que no merecen aprobar. Definido de esta forma el examen óptimo, para un estudiante que estudió UE unidades, se pueden plantear las siguientes posibles funciones objetivo particulares que deberán ser maximizadas:

- Probabilidad de la nota justa: El estudiante obtendrá la nota que merece. Esta es una medida difícil de satisfacer, ya que se debe asignar al estudiante una de las NM+1 categorías posibles (al asignarle una nota de 0 a NM); y es más difícil mientras mayor es NM:

$$f_1(UE) = P(NJ) \tag{9}$$

- Probabilidad de aprobación justa: El estudiante aprobará si merece aprobar, y desaprobará si no merece aprobar. Esta es una medida más relajada que la anterior debido a que considera sólo dos categorías, aprobado y desaprobado:

$$f_2(UE) = \begin{cases} 1 - PA & NJ < 4 \\ PA & NJ \geq 4 \end{cases} \tag{10}$$

Las funciones objetivo particulares planteadas miden qué tan óptimo es el examen tomado a un tipo particular de estudiante: el que estudió *UE* unidades. Para considerar a todos los estudiantes, se debe plantear una función objetivo global independiente de *UE*. Dos funciones objetivo globales posibles que cumplen con esa condición son las siguientes:

$$FO_1(UT) = \sum_{ue=0}^{UM} Pb(ue)\, f_1(ue) \tag{11}$$

$$FO_2(UT) = \sum_{ue=0}^{UM} Pb(ue)\, f_2(ue) \tag{12}$$

donde $Pb(ue)$ es la probabilidad de que un estudiante estudie *ue* unidades $P(UE = ue)$. Es decir, que las funciones objetivo globales planteadas son promedios probabilísticos de las funciones objetivo particulares analizadas anteriormente.

En este trabajo, se supone que $Pb(ue)$ obedece una distribución binomial:

$$Pb(ue) = \binom{UM}{ue} p^{ue} (1-p)^{UM-ue} \tag{13}$$

donde p es la probabilidad de que un estudiante estudie una unidad dada. Mientras más alta p, más dedicados son los estudiantes. Para el caso de estudio, se toma $p = 0.5$. La Fig. 1 presenta cómo varían las dos funciones objetivo globales para ese caso.

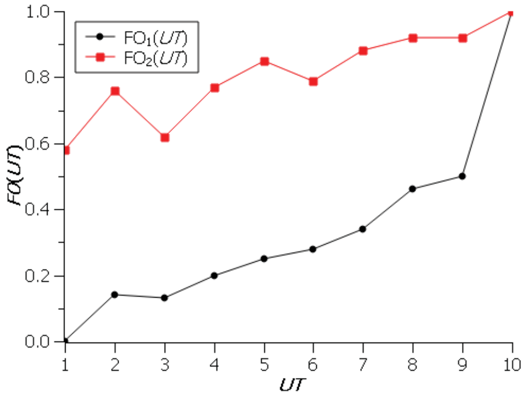

Fig. 1. Funciones objetivo globales el caso de estudio, con $NM = 10$, $NA = 4$, $UM = 10$ y $p = 0.5$.

Las funciones objetivo globales presentadas deben ser maximizadas para alcanzar un examen óptimo. Para ello, la única variable de decisión que dispone el profesor es *UT*. En la Fig. 1, se observa que los dos criterios planteados se cumplen cuando $UT = UM$. Sin embargo, esa solución tiene el mayor costo (duración del diseño del examen, duración del examen, duración de la corrección). Para que el costo de la evaluación no sea demasiado alto, es conveniente establecer la siguiente restricción:

$$UT \leq UT_{\max} \tag{14}$$

donde UT_{max} es la cantidad máxima de unidades que el profesor puede o está dispuesto a evaluar. Para el caso de estudio, se toma $UT_{max} = UM/2 = 5$.

En la Fig. 1, puede apreciarse que el valor de *UT* que cumple con la restricción planteada anteriormente y que maximiza ambas funciones objetivo es 5. Entonces, para el caso de estudio, lo recomendable es evaluar cinco unidades. En esa figura, también, puede apreciarse que al ser la segunda función objetivo más relajada que la primera, tiene valores más altos y es menos sensible a *UT*, lo que significa que es más probable que un examen sea óptimo si se realiza con la segunda función. Finalmente, es importante notar que, si bien ambas funciones objetivo presentan una tendencia creciente, no son monótonas crecientes; lo que hace posible que en algunos casos puedan existir soluciones no triviales al problema de optimización planteado.

4.1 Modelo de Optimización con Restricción UT

Para resolver el problema para cualquier *UM* y *p*, se planea el siguiente modelo de optimización INLP:

$$\begin{aligned}&\operatorname*{Max}_{UT} FO(UT)\\&\text{s. a:}\\&UT \le UT_{\max}\\&UT \in \mathbb{N}\end{aligned} \tag{15}$$

Debido a que la región factible es pequeña, este modelo se resolvió por búsqueda exhaustiva empleando una planilla Microsoft Excel.

La solución producida por este modelo garantiza el máximo valor de $FO(UT)$, pero no el mínimo valor de UT. Para el caso de existir múltiples soluciones, la solución deseada es la de UT mínimo. Para encontrar este valor, se resuelve un segundo problema de optimización INLP:

$$\begin{aligned}&\operatorname*{Min}_{UT} UT\\&\text{s. a:}\\&FO(UT) \ge FO_{\text{opt}}\\&UT \in \mathbb{N}\end{aligned} \tag{16}$$

donde FO_{opt} es el máximo valor de la función objetivo reportado por el primer problema de optimización.

La Fig. 2 presenta los resultados correspondientes a p = 0.5 y $UT_{\max} = UM/2$. Los valores $UT1_{\text{opt}}$ fueron obtenidos con $FO(UT) = FO_1(UT)$; mientras que los $UT2_{\text{opt}}$, con $FO(UT) = FO_2(UT)$. En dicha figura, puede apreciarse que la cantidad UT no siempre alcanza el máximo valor permitido por la restricción, con lo que se obtiene el resultado inesperado que a veces menos preguntas produce una mejor evaluación (llegando al extremo de 3 unidades evaluadas menos que la cantidad máxima permitida). Adicionalmente, al ser la segunda función objetivo más relajada que la primera, los valores óptimos de $UT2_{\text{opt}}$ son menores o iguales a $UT1_{\text{opt}}$, produciendo una evaluación de menor costo.

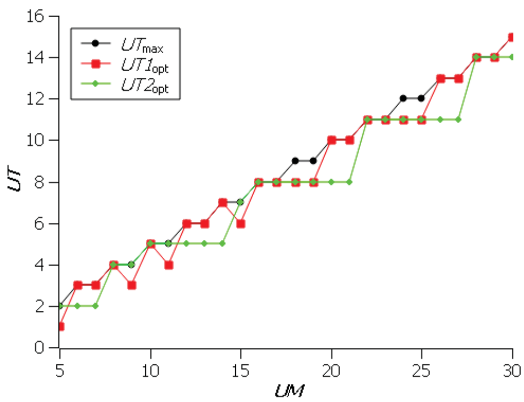

Fig. 2. Valores óptimos de UT para FO_1 y FO_2 con $NM = 10$, $NA = 4$ y $p = 0.5$.

4.2 Modelo de Optimización con Tolerancia

En la Fig. 1 se puede apreciar que no existe una gran diferencia entre $FO_1(2)$ y $FO_1(5)$, ni tampoco entre $FO_2(2)$ y $FO_2(5)$. Debido a esto, si el profesor tiene una cierta tolerancia, la nueva cantidad práctica de unidades a evaluar UT_{pra} puede ser menor a la UT_{opt} recomendada en la sección anterior.

Con este nuevo parámetro de tolerancia, el problema de optimización se descompone en dos problemas secuenciales. Primero, se debe resolver el primer problema de optimización planteado en la sección anterior, que determina el valor máximo de la función objetivo FO_{opt}. Luego, con ese valor, se debe resolver el siguiente problema de optimización INLP:

$$\begin{aligned} &\operatorname*{Min}_{UT} UT \\ &\text{s. a:} \\ &FO(UT) \geq FO_{\text{opt}} - Tol \\ &UT \in \mathbb{N} \end{aligned} \tag{17}$$

donde $Tol \in [0, 1]$ es la tolerancia, o merma en la probabilidad de realizar un examen óptimo, que acepta el docente.

Aplicando este modelo al caso de estudio ($NM = 10$, $NA = 4$, $UM = 10$, $NJ = NE$, $p = 0.5$ y $UT_{max} = 5$) con $Tol = 0.1$, se obtiene que la cantidad práctica de unidades a evaluar es 2 en lugar de 5 para la segunda función objetivo global, lo que es una apreciable reducción del costo de la evaluación.

La Fig. 3 y la Fig. 4 presentan los resultados para valores de Tol igual a 0.1 y 0.2, respectivamente, con $p = 0.5$ y $UT_{max} = UM/2$. Los valores $UT1_{pra}$ fueron obtenidos con $FO(UT) = FO_1(UT)$; mientras que los $UT2_{pra}$, con $FO(UT) = FO_2(UT)$. En dichas figuras, puede apreciarse que la cantidad UT está muy por debajo de la cantidad permitida y de la óptima, y esto se logra sin afectar demasiado la calidad de la evaluación.

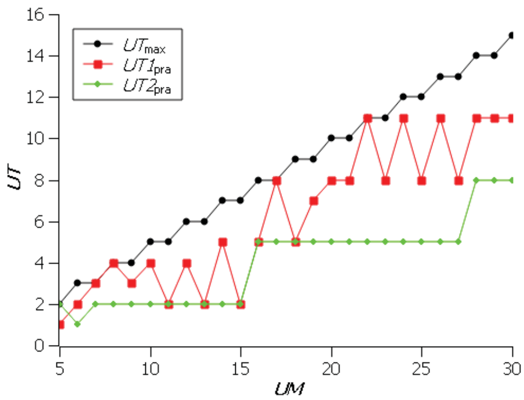

Fig. 3. Valores prácticos de UT para FO_1 y FO_2 con $NM = 10$, $NA = 4$, $p = 0.5$ y $Tol = 0.1$.

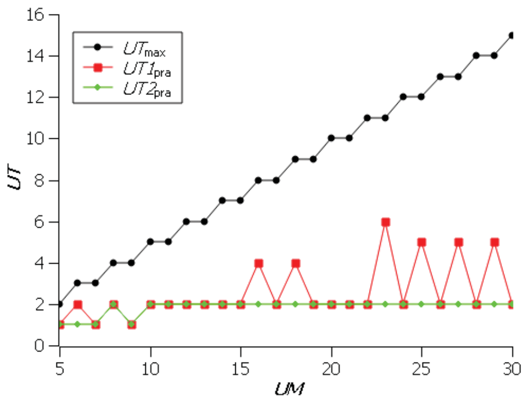

Fig. 4. Valores prácticos de UT para FO_1 y FO_2 con $NM = 10$, $NA = 4$, $p = 0.5$ y $Tol = 0.2$.

4.3 Efecto de la Nota de Aprobación

En el estudio realizado, se supuso que la nota de aprobación NA era igual a 4, que es el estándar de aprobación en algunas universidades. Sin embargo, otras universidades, o las cátedras, adoptan diferentes notas de aprobación de acuerdo a la instancia que están evaluando. Por ejemplo, una cátedra puede adoptar la nota 5 para aprobar un parcial, 7 para promocionar y 4 para aprobar el final. Teniendo en cuenta esta situación, se realizaron pruebas para analizar el efecto que tiene modificar NA. La modificación de la nota de aprobación modifica a $FO_2(UT)$ y, por lo tanto, también modifica a $UT2_{opt}$ y $UT2_{pra}$. No obstante, en las pruebas realizadas, si bien se observan cambios en $UT2_{pra}$ para $Tol = 0.2$ cuando se varía NA de 4 a 7, la diferencia es a lo sumo una unidad de evaluación.

4.4 Efecto de la escala

Cuando se incrementa la nota máxima *NM* a 100 y *NA* a 40, se observa un marcado deterioro de $FO_1(UT)$, mientras que $FO_2(UT)$ permanece casi inalterada. Este deterioro se debe a que, al aumentar la escala, las notas posibles también aumentan, y es menos probable que la nota del examen concuerde con la nota justa *NJ*. Por lo planteado, tanto $UT1_{pra}$ como $UT2_{pra}$ adoptan valores menores o iguales a 2 con $UT_{max} = 5$ y $Tol = 0.2$.

5 Conclusiones

En este trabajo, se definió un examen óptimo, y en base a esta definición, se formuló un modelo de optimización que determina la cantidad de unidades a evaluar que maximiza la probabilidad de tomar un examen óptimo. Este modelo se resolvió para diferentes casos, y se encontró que la cantidad óptima de unidades a evaluar no siempre fue la máxima permitida. Luego, se planteó un segundo modelo de optimización que permitió resolver el problema con una cierta tolerancia. Al aumentar la tolerancia que acepta el profesor, la cantidad de unidades a evaluar se redujo notablemente. También, se analizaron los efectos de modificar la mínima nota que necesita el estudiante para aprobar, y se encontró que no tiene un efecto importante sobre la cantidad de unidades que deben evaluarse para lograr un examen óptimo. Por otra parte, modificar la nota máxima del examen reduce notablemente la probabilidad de asignar una nota justa al estudiante.

Finalmente, del estudio realizado, se concluye que a veces menos preguntas producen una mejor evaluación, y que es más probable realizar un examen óptimo cuando se califica como aprobado o desaprobado, en lugar de asignar una nota.

Referencias

1. Frola, P., Velásquez, J.: Competencias docentes para la evaluación cuantitativa del aprendizaje. Centro de Investigación Educativa y Capacitación Institucional, D. F. México (2011)
2. Trillo Alonso, F., Porto Currás, M.: La percepción de los estudiantes sobre su evaluación en la universidad. Un estudio en la facultad de ciencias de la educación. Innovación Educativa, 9, 55-75 (1999)
3. Álamo Serrano, J.: Nuevas posibilidades de evaluación usando las TIC's: un vistazo a cuatro casos. En: La evaluación de los estudiantes en la Educación Superior, Servei de Formació Permanent, 54-73. Universidad de Valencia, Valencia (2007)
4. González-Santander, J.L., Martín, G.: Análisis de la fórmula para la calificación de pruebas tipo test multi-respuesta. Nereis. Revista Iberoamericana Interdisciplinar de Métodos, Modelización y Simulación, 3, 53-59 (2010)
5. Huapaya, C.R., Lizarralde, F.A., Arona, G.M.: Modelo basado en Lógica Difusa para el Diagnóstico Cognitivo del Estudiante. Formación Universitaria, 5(1), 13-20 (2012).
6. Bojadziev, G.: Fuzzy Sets, Fuzzy Logic, Applications (Series on Advances in Mathematics for Applied Sciences). World Scientific Publishing Company, New York (1996)
7. Walck C.: Hand-book on STATISTICAL DISTRIBUTIONS for experimentalists, University of Stockholm, Stockholm (2007)