FIABILIDAD DE UNA JERARQUÍA PARA EVALUAR EL RAZONAMIENTO PROBABILÍSTICO ACERCA DE LA DISTRIBUCIÓN BINOMIAL

Sánchez, E., Landín, P. R.

Departamento de Matemática Educativa

Cinvestav-IPN, México

Resumen. *Se describe un proceso para mejorar la fiabilidad de una jerarquía de razonamiento para evaluar las respuestas de estudiantes de bachillerato a tareas de distribución binomial. Se aplicó la jerarquía para realizar 6 evaluaciones independientes a un corpus de respuestas a un cuestionario con problemas de distribución binomial. La comparación y el análisis de las clasificaciones, en especial en las que no hubo consenso, llevaron a reestructurar la jerarquía. En una segunda clasificación, utilizando la jerarquía reestructurada, hubo un ligero aumento en la frecuencia de clasificaciones con consenso; no obstante, en conjunto hubo mayor homogeneidad de criterios que en la primera clasificación.*

Palabras clave: Jerarquía, fiabilidad, razonamiento probabilístico, distribución binomial.

Abstract. *The process followed to increase reliability of a reasoning hierarchy to evaluate the responses of high school students to tasks related to the binomial distribution is described. The process was to apply the hierarchy to perform 6 independent evaluations on a corpus of responses to a questionnaire with problems related to the binomial distribution. The comparison and analysis of the classifications, especially in the responses in that there was no consensus classification led to restructure the hierarchy. In a second classification, using the restructured hierarchy, there was a slight increase in the frequency of ratings consensus; however, there was more uniform set of criteria than the first classification.*

Key words: Hierarchy, reliability, probabilistic reasoning, binomial distribution.

INTRODUCCIÓN

La evaluación permite observar en qué medida se cumplen los objetivos institucionales y verificar si estos efectivamente son alcanzados. Su función es también informar al profesor sobre los efectos de su enseñanza, además de ser un indicador del aprovechamiento de sus estudiantes. Las jerarquías de razonamiento son instrumentos útiles para la evaluación. Reading y Reid (2010) sostienen que las jerarquías: 1) apoyan el análisis y el desarrollo del currículo desde un punto de vista cognitivo; 2) apoyan el diseño y elaboración inicial de secuencias de aprendizaje, así como su transformación y adaptación a las necesidades de los estudiantes; 3) ayudan a seleccionar tareas de evaluación del desarrollo cognitivo, y 4) permiten al profesor apreciar las relaciones entre conceptos y fomentarlas entre sus estudiantes.

Sobre el uso jerarquías, Tarr y Jones (1997) presentan una acerca de las nociones de probabilidad condicional e independencia; Jones, Thornton, Langrall y Tarr (1999) la extienden además a tareas de espacio muestral, probabilidad teórica, probabilidad experimental y comparación de probabilidades; Watson y Moritz (1999a) elaboran una para tareas de comparación entre gráficas de datos; Watson y Moritz (1999b) para la noción de promedio; Reading y Shaughnessy (2004) y Reading y Reid (2007) para la variación, mientras que Aoyama (2007) para la interpretación de gráficas al igual que las categorías de Arteaga y Batanero (2010).

Reading y Reid (2006) ofrecen una jerarquía útil para distribuciones de datos empíricos, sin embargo, en la literatura no se encontraron jerarquías para describir el razonamiento sobre distribuciones teóricas específicas. En Landín y Sánchez (2010) se ha propuesto una jerarquía acerca de la distribución binomial, en esta comunicación se describe el proceso realizado para aumentar su fiabilidad.

PLANTEAMIENTO DEL PROBLEMA

La jerarquía objeto de este estudio fue construida para evaluar el razonamiento probabilístico de los estudiantes acerca de y con la distribución binomial con base en las respuestas a tareas sobre el tema. Como instrumento de evaluación conviene que sea *fiable*, es decir, que se obtengan los mismos resultados, con prudente margen de error, cuando se hagan dos o más evaluaciones independientes al mismo *corpus*. Por la complejidad del razonamiento probabilístico y la variedad potencial de detalles en las respuestas de los estudiantes a tareas sobre un tema, se asume que no es posible asegurar una total fiabilidad de una jerarquía para evaluar dicho razonamiento. No obstante, es posible mejorar su aplicabilidad mediante la búsqueda de un consenso subjetivo. Esta técnica consiste en someter el mismo material a la evaluación de varios participantes competentes utilizando la jerarquía en estudio; ésta se modifica buscando un acuerdo con base en la observación y análisis de las diferencias. En la presente comunicación se informa sobre una solución al problema de mejorar la fiabilidad de una jerarquía para evaluar el razonamiento probabilístico sobre la distribución binomial.

METODOLOGÍA

Participaron seis profesionales competentes en educación matemática (incluyendo a los autores) y 66 estudiantes de bachillerato (17-18 años). Se utilizaron un cuestionario de probabilidad de 4 preguntas (ver Apéndice) y las respuestas de los 66 estudiantes que se clasificaron con base en la jerarquía de la tabla 1. Se siguió el procedimiento descrito en seguida:

1) Se ofreció un taller de 10 sesiones a los 66 estudiantes divididos en 2 grupos. En él se resolvieron problemas de tipo binomial con ayuda del software Fathom.

2) Al cabo del taller, los estudiantes resolvieron el cuestionario en una sesión 1:30 horas. Se observaron y analizaron las respuestas de los estudiantes.

3) Los 6 evaluadores independientemente uno del otro clasificaron las respuestas con base en la jerarquía.

4) Se reunieron las clasificaciones y se observaron las coincidencias y diferencias.

5) Se discutieron entre los 6 evaluadores los motivos de las divergencias y se acordaron ajustes para reestructurar la jerarquía (cuadros 3 y 4).

6) Los seis participantes volvieron a clasificar las respuestas con base en la jerarquía reestructurada.

7) Se observó el nivel de coincidencia alcanzada en esta segunda clasificación y se comparó con la primera clasificación.

Nivel 1: subjetivo	**Nivel 2: transicional**	**Nivel 3: cuantitativo informal**	**Nivel 4: numérico**
• Aunque se reconoce la distribución de Bernoulli el razonamiento es idiosincrásico o basado en sesgos cognitivos; por ejemplo, al evaluar probabilidades binomiales se presenta el sesgo de representatividad, el de equiprobabilidad o se cae en la *ilusión de la linealidad*[81].	• Se representan los elementos, no necesariamente de manera exhaustiva, mediante secuencias de E^s (éxitos) y F^s (fracasos). Se considera el parámetro *n* de la binomial en la longitud de las secuencias. • Para evaluar la probabilidad se utiliza, bien o mal, la definición clásica (de Laplace) de probabilidad o se utiliza el valor esperado. Revierte frecuentemente a razonamiento subjetivo.	• Se reconoce el carácter combinatorio de la situación. Se considera la variable *k* al separar los casos favorables. • El cálculo de combinaciones se presenta en dos niveles: - mediante una lista organizada o - utilizando diagrama de árbol para calcular los casos favorables • Se aplica la definición clásica o, a veces, se aplica la regla del producto para calcular la probabilidad de una secuencia.	• Se reconocen la variable aleatoria y sus parámetros; su carácter combinatorio y se aplica la regla del producto. Se calculan las combinaciones apoyándose en el triángulo de Pascal o en la fórmula. • O se plantea la solución en términos de una instancia de la expresión general de la distribución binomial B(*n*, *p*, *k*) o del binomio de Newton. Se obtiene la solución mediante el uso de tablas, calculadora, computadora o mediante cálculo directo en la fórmula

Tabla 1. Jerarquía para evaluar el razonamiento probabilístico de y con la distribución binomial

RESULTADOS

Después de que los 6 evaluadores hicieron su clasificación, se contó el número de respuestas en las que hubo consenso entre los evaluadores sobre los niveles en que se clasificaron las respuestas (137 que representa 52%) y en las que hubo diferencias (121, es decir 46%); en la tabla 2, se puede observar cómo se distribuyeron por pregunta.

[81] La ilusión de la linealidad, de acuerdo con Van Dooren, et al. (2003), es una fuerte tendencia de los estudiantes a resolver problemas, con relaciones proporcionales aparentes, mediante la proporcionalidad o propiedades de las funciones lineales.

Concepto	Problema 1	Problema 2	Problema 3	Problema 4	Total	Total (%)
Consenso	32	48	43	14	137	52%
Sin consenso	32	18	21	50	121	46%
No precisados	2	0	2	2	6	2%
TOTAL	66	66	66	66	264	100%

Tabla 2. Niveles de consenso en la 1a clasificación

Se analizaron las razones que cada evaluador arguyó para hacer sus asignaciones en las preguntas en las que no hubo consenso; como resultado se identificaron ambigüedades u omisiones en la jerarquía utilizada en la primera clasificación (tabla 1). Se encontró que las divergencias tienen su origen en que la jerarquía que se utilizó en la primera clasificación sólo establece reglas de decisión basadas en la identificación de rasgos que indican si el estudiante conoce tal o cual elemento de conocimiento, pero es neutral sobre la precisión con que se aplican o sobre la corrección de la ejecución. Así, llamamos *elementos de ejecución* a las características enumeradas en la siguiente lista, mismas que se tomaron en cuenta para reestructurar la jerarquía:

- Sin cálculos explícitos
- Errores de cálculo
- Respuestas parciales
- Inconsistencias o con componentes extrañas

En seguida se ilustra con ejemplos cómo emergieron estos elementos:

Las diferencias en la clasificación a algunas respuestas se deben a la observación de que en ellas no se presentan *cálculos explícitos,* por ejemplo, la siguiente respuesta al problema 1:

[Responde inciso c, son igualmente probables] *Porque es difícil que salga HHHHH y que salga HMHMM en ese orden así que su probabilidad es la misma (c)*

Dos evaluadores la clasificaron en T (Transicional) y cuatro en S (Subjetivo). Los que clasificaron en T observaron que el estudiante era consciente de que las representaciones forman parte de un conjunto ordenado de secuencias, manifestado en la expresión "en ese orden", aparte de haber elegido la respuesta correcta. Uno de los que la clasificó como S comentó "aunque es la respuesta correcta no realiza cálculo".

Otras discrepancias provienen de la omisión de reglas de decisión referentes a *errores de cálculo* en la jerarquía; un ejemplo es la siguiente respuesta al problema 2,

[El estudiante eligió como respuesta (incorrectamente) el inciso b]. *$(X_2=1)=2(.5)(.5)=.045$; $(X_4 = 2) = 6\ (.5)(.5)(.5)(.5) =0.375$; Es más probable obtener 2 águilas en 4 volados.*

Dos evaluadores la clasificaron como N (Numérica), uno dudó entre N o CI (Cuantitativo Informal) y 2 la clasificaron en T. Los primeros vieron que en la respuesta se refleja que el estudiante conoce el coeficiente binomial y la regla del producto; características que en la descripción de la jerarquía llevan a la categoría N. El evaluador que dudó entre N o CI consideró además el hecho de que el resultado no es correcto por errores en los cálculos, esto le hizo pensar que debía clasificarse en un nivel menor, pero dudó al no estar explícito dicho criterio en la jerarquía. Otros la clasificaron en T, pues consideraron los errores en el cálculo.

La respuesta al problema 4, que se muestra en seguida, produjo diferentes reacciones de los evaluadores:

$$P(X=3)=C_3^5(70)^3(30)^{5-3}=\frac{5!}{3(2!)}=.3087\ .$$

Un evaluador la clasificó como N, otro dudó entre CI o N, dos en CI y uno en T. Los primeros vieron que el segundo miembro claramente muestra que se conoce la fórmula de la distribución binomial y el tercer miembro que se conoce la fórmula para calcular los coeficientes binomiales, además 0.3087 es el valor correcto para X=3; la jerarquía indica que con tales características la respuesta debe pertenecer a N. El evaluador que duda entre N o CI hizo la misma observación, pero bajó una categoría debido a los errores que son evidentes. Los que lo clasificaron en T, notaron que se respondía solamente a una parte del problema, pues falta calcular P(X=4) y P(X=5), además se expresaron inadecuadamente las probabilidades y las igualdades no son verdaderas.

Consecuencias del análisis de las clasificaciones sin consenso

Las clasificaciones de las respuestas y opiniones de los evaluadores pusieron de manifiesto que una fuente de ambigüedad fue la ausencia en la jerarquía de indicaciones que tuvieran en cuenta *elementos de ejecución* (cálculos explícitos, errores de cálculo, respuestas parciales, coherencia y componentes extrañas). Algunos evaluadores se sintieron incómodos al clasificar dos respuestas en un mismo nivel sólo porque en la respuesta analizada se podía percibir que los estudiantes conocían los conceptos previstos en la jerarquía (por ejemplo regla del producto) pero la ejecución en un caso era correcta y en la otra no; otros tomaron decisiones de incorporar estos elementos de ejecución como criterios para su clasificación. Por lo demás, parece natural evaluar el nivel de conocimiento de un concepto no sólo por su mención o manifestación sino también por la manera en que se utiliza. La misma razón influyó en la percepción de la necesidad de incorporar un nuevo nivel de razonamiento A (abstracto) que diera un lugar a aquellas respuestas que mostraban que el estudiante tenía conocimiento de las componentes de razonamiento, previstas en la jerarquía, y además los aplicaba y organizaba de manera precisa.

En la tabla 3, se muestra la técnica que se utilizó para incorporar los elementos de ejecución en la jerarquía. Se procedió enlistando en la primera fila las componentes de razonamiento y en la primera columna elementos de ejecución. Estos fueron ordenados de la componente de ejecución más grave a la más adecuada. Después se hizo la clasificación para la fila de respuestas correctas (la última fila de la tabla) de la misma manera que en la jerarquía definida arriba. A partir de esto se convino ponderar la presencia de fallas en la ejecución asignando niveles inferiores. La asignación es arbitraria pero permite que se refleje en la clasificación la calidad de las respuestas.

	Respuestas idiosincrásicas	Representación de las secuencias y probabilidad clásica	Regla del producto	Coeficiente binomial	Fórmula binomial
Sin coherencia o con componentes extrañas	S	S	S	S	S
Sin algunos cálculos explícitos	S	S	T	T	CI
Errores de cálculo	S	S	T	CI	N
Respuestas parciales	S	T	CI	N	N
Respuesta correcta s/defectos anteriores	S	CI	N	N	A

Tabla 3. Guía para incorporar los elementos de ejecución en la jerarquía

Nivel 1. Subjetivo.

• Aunque se reconoce la distribución de Bernoulli y posiblemente se elija la opción correcta no la argumenta en absoluto o la argumentación es de tipo idiosincrásico y/o con base en algún sesgo cognitivo.

• Es posible que se muestren rasgos de conocimientos pertinentes, como diagramas de árbol o la regla del producto pero sin coherencia, con componentes extraños o con gran cantidad de errores

Nivel 2. Transicional

• Se representan los elementos del Espacio Muestral de manera conveniente aunque no necesariamente de manera exhaustiva.

• Así mismo, para evaluar la probabilidad, se utiliza, bien o mal, la definición clásica de probabilidad o se usa el valor esperado.

• También se clasifican en esta categoría respuestas que se apoyan en la regla del producto y/o calculan o usan coeficientes binomiales, pero de manera parcial, o sin hacer los cálculos correspondientes o con errores importantes

Nivel 3. Cuantitativo Informal

• Se reconocen el carácter combinatorio de la situación. Se considera la variable k al separar los casos favorables.

• Se calculan combinaciones mediante una lista organizada o mediante la elaboración un diagrama de árbol

• Se aplica la definición clásica de probabilidad y se llega a la respuesta correcta.

• De manera alternativa, se aplica la regla del producto aunque con resultados parcialmente correctos.

• En ocasiones se calculan y usan los coeficientes binomiales pero con errores de cálculo o se utiliza la fórmula binomial pero se omiten los cálculos

Nivel 4. Numérico

• Se reconocen la variable aleatoria, los parámetros de la distribución binomial y el carácter combinatorio de la situación. Se utiliza la regla del producto y se cuentan las combinaciones

para obtener una respuesta correcta. • La respuesta puede ser parcialmente correcta, pero utiliza la regla del producto y calcula el coeficiente binomial con el triángulo de Pascal o con la fórmula de las combinaciones o, también puede clasificarse en esta categoría respuestas que hacen uso de la fórmula de la distribución binomial, aunque con algunos errores o con resultados parciales
Nivel 5. Abstracto
• Se reconocen los parámetros de una distribución binomial; se plantea el problema y obtiene la solución en términos de una instancia de la expresión general de la distribución binomial B(n, p, k) o del binomio de Newton. • Se reconocen la influencia de los parámetros n y p en la forma gráfica de la distribución

Tabla 4. Jerarquía para el razonamiento binomial, reestructurada a partir de la del tabla 1.

En la tabla 4 se presenta la descripción de la jerarquía reestructurada considerando los elementos de ejecución y las convenciones expresadas en la tabla 3. Después de aplicar esta nueva jerarquía el nivel de consenso se incrementó en 11%.

CONCLUSIONES

La jerarquía resultante proporciona una forma de evaluar el razonamiento probabilístico de los estudiantes acerca de la distribución binomial y es útil para clasificar las respuestas a las preguntas del cuestionario (ver Apéndice); sin embargo, el hecho de que haya sido construida con base en los elementos que componen la distribución binomial y en aspectos de la ejecución hacen que la jerarquía sea aplicable a más tareas acerca de la distribución binomial. A partir de esto, se puede extender su función a los puntos señalados por Reading y Reid (2010) relacionados con: 1) aspectos cognitivos del currículo; 2) diseño de secuencias de aprendizaje; 3) selección de tareas y 4) guía para el profesor.

En este estudio también se muestra que se puede mejorar la fiabilidad de la jerarquía mediante la búsqueda de un consenso intersubjetivo, es decir, la discusión de cómo se aplica por los participantes y de la solución de las ambigüedades que emergen en las divergencias en la clasificación. Tales soluciones pueden traducirse en la introducción de nuevos criterios en la jerarquía para capturar diferencias importantes. Sin embargo, no se debe tener la pretensión de querer prever y dar cuenta de todos los rasgos potencialmente plausibles de aparecer en las respuestas de los estudiantes, pues de esa manera la jerarquía puede volverse inmanejable; para que sea útil debe mantener un cierto nivel de generalidad.

A pesar de que el incremento del consenso fue modesto, los evaluadores consideraron que se había reducido considerablemente la incertidumbre que sintieron con la primera jerarquía a la hora de decidir. De cualquier manera, lo importante es que si un equipo utiliza la jerarquía la adecúe a las circunstancias particulares de su investigación o de sus clases.

Agradecimientos

Este trabajo es financiado por CONACYT, proyecto 101708.

Agradecemos a Silvia Mayén, Antonio Orta, Julio César Valdés, Jaime García, Fátima Rubiales y María Guzmán su colaboración.

Referencias

Aoyama, K. (2007). Investigating a hierarchy of students' interpretations of graphs. *International Electronic Journal of Mathematics Education*, *4*(3), 298-318.

Recuperado el 22/10/2010 de: www.iejme.com

Arteaga, P. y Batanero, C. (2010). Evaluación de errores de futuros profesores en la construcción de gráficos estadísticos. En M. Moreno, A. Estrada, J. Carrillo, T.A. Sierra (Eds.), *Investigación en Educación Matemática XIV (210-221).* Lleida, España: SEIEM y Universidad de Lleida.

Jones, G.A., Thornton, C.A., Langrall, C.W. y Tarr, J.E. (1999). Understanding student's probabilistic reasoning. En L.V.Stiff & F.R.Curcio (Eds.), *Developing mathematical reasoning in grades K-12* (1999 yearbook, pp. 146-155). Reston, VA: National Council of Teachers of Mathematics.

Landín, P. R. y Sánchez, E. (2010). Niveles de razonamiento probabilístico de estudiantes de bachillerato frente a tareas de distribución binomial. *Educação Matemática Pesquisa*, *12* (3). [Recuperado el 01/03/11 de:

http://revistas.pucsp.br/index.php/emp/article/viewArticle/4842]

Reading, C. y Reid, J. (2006). An emerging hierarchy of reasoning about distribution: from a variation perspective. *Statistics Education Research Journal*, *5*(2), 46-68.

Recuperado el 27/10/2009 de: http://www.stat.auckland.ac.nz/serj.

Reading, C. y Reid, J. (2007). Reasoning about variation: Student voice. *International Electronic Journal of Mathematics Education*, *2* (3), 110-127.

Reading, C. y Reid, J. (2010). Reasoning about variation: rethinking theoretical frameworks to inform practice. En C. Reading (Ed.*), Data and context in statistics education: Towards an evidence-based society*. *Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8)*, July, 2010. Ljubljana, Slovenia. Voorburg, The Netherlands: International Statistical Institute.

Reading, C. y Shaughnessy, J. M. (2004). Reasoning about variation. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (201-226). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Tarr, J. E., y Jones, G. A. (1997). A framework for assessing middle school students' thinking in conditional probability and independence. *Mathematics Education Research Journal*, *9*(1), 39-59.

Van Dooren, W., De Bock, D., Depaepe, F., Janssens, D. y Verschaffel, L. (2003). The illusion of linearity: expanding the evidence towards probabilistic reasoning. *Educational Studies in Mathematics*, *53*(2), 113-138.

Watson, J. M. y Moritz, J. B. (1999a). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics*, *37*(2), 145-168.

Watson, J. M. y Moritz, J. B. (1999b). The development of the concept of average. *Focus on Learning Problems in Mathematics*, *21*(4), 15-39.

APÉNDICE: Cuestionario

1. Considere todas las familias en las que hay 5 hijos. Se elige una familia al azar. Se supone que la probabilidad de ser mujer es igual a la de ser hombre y es igual a ½. De las siguientes afirmaciones cuál es correcta:

   a) El evento HMHMM es más probable que el evento HHHHH

   b) El evento HHHHH es más probable que el evento HMHMM

   c) El evento HHHHH es igual de probable que el evento HMHMM

   Donde HMHMM significa que el mayor es Hombre, el que sigue Mujer, después Hombre y los dos menores Mujeres. HHHHH significa que todos son Hombres.

   Justifica tu respuesta:

2. ¿Qué es más probable?:

   a) obtener 1 águila en 2 volados

   b) obtener 2 águilas en 4 volados

   c) son igualmente probables

   Justifica tu respuesta:

3. En un hospital hay dos médicos; uno atiende 3 partos al día, mientras que el otro atiende 6 partos al día. Un día hacen una apuesta: observan el número de niños varones que nacen en los partos que cada quien atiende y gana quien en los partos que atiende iguale o rebase el 60% de niños varones (se supone que la probabilidad de que el recién nacido sea varón es igual a la probabilidad de que sea mujer). ¿Quién tiene mayor probabilidad de ganar?

   a) El médico que atiende 3 partos tiene más probabilidad de ganar

   b) El médico que atiende 6 partos tiene más probabilidad de ganar

   c) Ambos tienen la misma probabilidad

   Justifica tu respuesta

4. Un tirador tiene 70% de probabilidad de pegar en el blanco y 30% de fallar. En un concurso gana si le pega al blanco al menos 3 veces de 5. ¿Cuál es su probabilidad de ganar?