

Sesgos en el Razonamiento Sobre Probabilidad Condicional e Implicaciones Para la Enseñanza

Carmen Batanero.

batanero@ugr.es

Universidad de Granada

España

J. Miguel Contreras.

jmcontreras@ugr.es

Universidad de Granada

España

Carmen Díaz.

carmen.diaz@dpsi.uhu.es

Universidad de Huelva

España

Resumen.

En este trabajo analizamos algunos de los sesgos frecuentes en la comprensión de la probabilidad condicional. También presentamos algunos resultados obtenidos en dos estudios de evaluación realizados con estudiantes de Psicología y futuros profesores de matemáticas en España. Finalizamos con algunas reflexiones sobre la enseñanza de la probabilidad.

Palabras clave: Probabilidad condicional, sesgos, evaluación

Abstract.

In this paper we analyze some biases that are frequent in understanding conditional probability. We also present results from two assessment studies carried out with Psychology students and prospective secondary school mathematics teachers in Spain. We conclude with some reflections on the teaching of probability.

KeyWords: Conditional probability, biases, assesment

1.1 Introducción

La probabilidad condicional es un concepto base al aplicar la Estadística, porque permite incorporar cambios en nuestro grado de creencia sobre los sucesos aleatorios, a medida que adquirimos nueva información. Es también un concepto teórico requerido en la construcción del espacio muestral producto. Por ello, su correcta comprensión y el razonamiento sobre la misma son requisitos en el estudio de la inferencia estadística, tanto clásica como bayesiana, así como en el estudio de la asociación entre variables, la regresión y los modelos lineales. En el terreno profesional y la vida cotidiana, la toma de decisiones acertadas en situaciones de incertidumbre se basa en gran medida en el razonamiento condicional. Todo ello hace que el tema se incluya tanto en la educación secundaria, como en la universidad en muchos países.

A pesar de que este tema se enseña en estos dos niveles educativos, la investigación psicológica y didáctica sugiere la existencia de intuiciones incorrectas, sesgos de razonamiento y errores de comprensión y aplicación de este

concepto. En lo que sigue analizamos algunas de estas investigaciones y mostramos los resultados obtenidos en dos investigaciones propias, donde se aplicaron diversos ítems que evalúan la presencia de estos sesgos.

En la primera de estas investigaciones [1] participaron 414 estudiantes de las universidades de Granada (cuatro grupos de estudiantes; $n = 307$), y Murcia (dos grupos; $n = 106$). Todos los participantes cursaban Primer Curso de Psicología y eran alumnos de la Asignatura de Análisis de Datos. La mayoría de los alumnos tenían una edad de 18 o 19 años. Los estudiantes habían estudiado el tema durante dos semanas, aproximadamente un mes antes que se pasase el cuestionario. Se realizó la toma de datos después del examen parcial que incluía el tema, para asegurarse que los alumnos lo habían estudiado.

En España el acceso a profesor de matemáticas de secundaria se realiza mediante un concurso. Para poder realizar este concurso, actualmente se exige un título específico de Máster Universitario en Formación del Profesorado de Educación Secundaria Obligatoria y Bachillerato, Formación Profesional y Enseñanza de Idiomas, en la especialidad de matemáticas (en lo sucesivo Máster de Secundaria). Aproximadamente el 50% de los que acceden a dicho Máster son Licenciados de Matemáticas y el resto tienen diversas titulaciones científicas o técnicas. Por otro lado, dentro de la Licenciatura de Matemáticas, el 90% de los egresados realizan el concurso para profesor de matemáticas de secundaria, al finalizar su formación. En consecuencia, los profesores de matemática en España provienen en la misma proporción de dos tipos de estudios: (a) o bien son licenciados en Matemáticas (el 90% de dichos titulados) o bien son egresados del Máster de Secundaria. Para conseguir una muestra representativa de futuros profesores de secundaria españoles en el segundo estudio [2] se decidió tomar estos dos tipos de alumnos, eligiendo varias universidades para lograr un tamaño de muestra adecuado y mejorar la representatividad, puesto que el número de alumnos, tanto en la licenciatura de matemáticas, como en el Máster es pequeño en cada universidad. Participaron 196 futuros profesores de matemáticas de educación secundaria (95 alumnos de último curso de la licenciatura de Matemáticas de las Universidades de Granada, La Laguna y Salamanca y 101 alumnos del Máster de Secundaria, de las Universidades de Alicante, Barcelona, Cádiz, Extremadura, Granada, Salamanca, Santiago de Compostela, Pública de Navarra y Valladolid). El rango de edad fue de 22 a 30 años.

La finalidad de este trabajo es analizar algunas dificultades con las que nos enfrentamos al razonar sobre la probabilidad condicional, mostrando ejemplos de tareas que permiten evaluarlas, así como las respuestas típicas a las mismas. Creemos que esta información es valiosa para la planificación de la enseñanza de la probabilidad y la evaluación del aprendizaje.

1.2 Definición de la Probabilidad Condicional e Independencia

La probabilidad condicional puede definirse con diversos grados de formalización. Intuitivamente podemos decir que la probabilidad condicional $P(A/B)$ de un suceso A dado otro suceso B es la probabilidad de que ocurra A sabiendo que B se ha verificado. Desde un punto de vista más formal, se define mediante la expresión:

$$P(A/B) = P(A \cap B)/P(B), \text{ siempre que } P(B) > 0.$$

El primer punto en nuestros estudios fue ver si estas definiciones se comprenden, para lo cual se pidió a los alumnos que dieran de una definición intuitiva de probabilidad simple y de probabilidad condicional, aportando un ejemplo de cada una. Los resultados (Tabla 1) sugieren que, aunque una parte importante de los alumnos define correctamente las probabilidades, la tercera parte no da respuesta o tiene imprecisiones en la definición.

	Futuros profesores ($n = 196$)		Psicología ($n = 414$)	
	Frecuencia	%	Frecuencia	%
Define incorrectamente ambas probabilidades	33	16,9	119	28,7

Define imprecisamente una probabilidad	6	3,2	28	6,8
Define correctamente una probabilidad	12	6,0	90	21,7
Define imprecisamente ambas probabilidades	114	58,1	50	12,1
Define correctamente ambas probabilidades	31	15,8	127	30,7
Total	196	100,0	414	100,0

Tabla 1. Respuestas en la definición de la probabilidad simple y condicional.

Algunos participantes utilizaron expresiones imprecisas para definir algunas de las probabilidades, como por ejemplo: “En la probabilidad condicional, para que se dé un suceso, se tiene que dar otro”. Este alumno indica correctamente que en la probabilidad condicional intervienen dos sucesos, pero la respuesta es imprecisa porque matemáticamente podemos definir la probabilidad condicional, independientemente de que el suceso ocurra o no. Otro ejemplo de definición imprecisa sería: “Probabilidad simple: aquella en la que hay un sólo elemento y en la probabilidad condicional intervienen dos sucesos”. La respuesta es imprecisa, porque en la probabilidad conjunta también intervienen dos sucesos. La respuesta: “La probabilidad simple es la probabilidad de que ocurra una variable y la condicional que ocurra sabiendo que ha ocurrido otra que la condiciona” es imprecisa porque el alumno se refiere a variables y no a sucesos. Como se muestra en la tabla los resultados fueron peores en los futuros profesores de matemáticas que en los estudiantes de psicología.

Un concepto relacionado con la probabilidad condicional es el de independencia. Matemáticamente puede deducirse de la regla del producto de probabilidades, mediante la expresión:

$$A \text{ y } B \text{ son independientes si y sólo si } P(A \cap B) = P(A) \times P(B)$$

Este concepto se relaciona con el de probabilidad condicional, ya que dos sucesos son independientes si la probabilidad de uno de ellos no cambia al condicionarlo por el otro [3]. Aunque también en este caso la definición es simple, desde un punto de vista psicológico y didáctico es difícil, en muchos casos, saber si dos sucesos son o no independientes, al resolver un problema o al tomar una decisión. Maury analizó la comprensión intuitiva de la probabilidad condicional de 374 estudiantes de los últimos cursos de Bachillerato, usando cuatro problemas en un contexto de extraer bolas de urnas; con y sin reemplazamiento. También utilizó dos tipos de vocabulario (técnico y cotidiano) al plantear, observando que sólo la cuarta parte de los alumnos daban respuestas correctas en el cálculo de probabilidad condicional mientras que los mismos alumnos obtuvieron un 60% de aciertos en problemas de probabilidad simple.

Maury ([3]) supuso que la dificultad no está ligada, en sí, a la noción de independencia, sino al hecho de que los dos sucesos (azul /rojo) sean no equiprobables que introduce un distractor que aumenta la dificultad de la tarea. Esta suposición se confirmó en otro experimento [4] en la que plantea a 290 alumnos de entre 13 y 16 años un problema similar en el contexto de lanzamiento de una moneda, obteniendo en este caso un 70 % de éxitos, lo que para Maury indica el reconocimiento intuitivo de la independencia por parte de los alumnos. Nuestra versión del problema en un contexto de lanzamiento de dados es el siguiente:

Problema 1. Una persona lanza un dado y anota si saca un número par o impar. Se trata de un dado no sesgado (es decir todos los números tienen la misma probabilidad). Estos son los resultados al lanzarlo 15 veces: par, impar, impar, par, par, impar, par, par, par, par, impar, impar, par, par, par. Lanza el dado de nuevo ¿Cuál es la probabilidad de sacar un número par en esta nueva tirada?

En la Tabla 2 presentamos los resultados del problema 2, donde la mayoría de los alumnos da una respuesta correcta, siendo los resultados mucho mejores en los futuros profesores de matemáticas. Un 5% de futuros profesores y 20%

de estudiantes de Psicología obtiene una estimación frecuencial de la probabilidad, es decir, calculan la probabilidad a partir de la frecuencia relativa obtenida en los ensayos que se describen en el ítem, dando como valor de la probabilidad el valor 10/15 (hay un total de 10 pares en los quince lanzamientos). Un 16,8% de futuros profesores y 27% de estudiantes de psicología da respuestas son erróneas, que pueden clasificarse en tres categorías: (1) o bien se calcula la probabilidad de ocurrencia en función de los resultados anteriores, cometiendo la falacia del jugador, donde se espera que una serie corta de ensayos se equilibre [5] y [6]; (2) tratan de aplicar la fórmula de la probabilidad condicional, pero comete un error en la fórmula, (3) hacen una interpretación incorrecta del enunciado o no responden.

	Futuros profesores ($n = 196$)		Psicología ($n = 414$)	
	Frecuencia	%	Frecuencia	%
No responde o asume dependencia	33	16,8	112	27,0
Estimación frecuencial probabilidad	10	5,1	84	20,3
Respuesta correcta	153	78,1	218	52,7
Total	196	100,0	414	100,0

Tabla 2 Respuestas al problema 1

La dificultad del concepto de independencia no es exclusiva de los participantes en nuestro estudio pues Sánchez ([6]) pasa un cuestionario de probabilidad a 88 profesores de Matemáticas que participaban en México en un programa de actualización, proponiéndoles el problema 2, encuentra que sólo 44 de los profesores hicieron intentos sistemáticos por resolver el problema.

Problema 2. Se extrae una carta al azar de una baraja americana: sea A el evento “se extrae un trébol” y B el evento “se extrae una reina” ¿Los eventos A y B son independientes?

De los profesores que tratan de resolverlo, 39 llegaron a una respuesta, pero sólo 4 lo hicieron correctamente, utilizando la regla del producto. En las respuestas incorrectas encuentra dos tipos de razonamiento:

1. Creer que eventos independientes son lo mismo que eventos excluyentes: “No son independientes porque tenemos la reina de tréboles”. Este es un error muy extendido, y ha sido descrito, entre otros autores, por Kelly y Zwiers [5]. Los autores sugieren que el error puede ser debido a la imprecisión del lenguaje ordinario, en que “independiente” puede significar, a veces, separado. Por otro lado, tanto en la definición formal de independencia como en la de sucesos excluyentes interviene la operación de intersección. Usualmente decimos que dos sucesos son independientes cuando la aparición/no-aparición de uno de ellos no proporciona información sobre la ocurrencia del otro. En el caso de sucesos excluyentes la aparición de uno implica la no-aparición del otro, por tanto dos sucesos excluyentes son siempre dependientes.
2. Creer que sólo se puede aplicar la idea de independencia a sucesiones de experiencias: “Si extraemos una carta para verificar el evento A y se vuelve a colocar en la baraja para verificar el evento B entonces A y B son independientes. Si se extrae la carta para verificar A y no se regresa, entonces A y B no son independientes”.

En nuestra investigación propusimos este mismo problema, pero se incluyeron en las respuestas alternativas los errores descritos por Sánchez ([6]), quedando de la siguiente forma:

Problema 3. Se extrae una carta al azar de una baraja española (40 cartas con cuatro palos: oros, bastos, espadas y copas. Cada palo tiene los números del 1 al 7, sota caballo y rey). Sea A el suceso "se extrae una carta de oros" y B el suceso "se extrae un rey". ¿Los sucesos A y B son independientes?

1. No son independientes, porque en la baraja hay un rey de oros.
2. Sólo si sacamos primero una carta para ver si es rey y se vuelve a colocar en la baraja y luego sacamos una segunda para ver si es oros.
3. Sí, porque $P(\text{rey de oros}) = P(\text{rey}) \times P(\text{oros})$.
4. No, porque $P(\text{rey/oros}) \neq P(\text{rey})$.

	Futuros profesores ($n = 196$)		Psicología ($n = 414$)	
	Frecuencia	%	Frecuencia	%
a.) No son independientes porque en la baraja hay un rey de oros.	63	32,1	118	28,5
b.) Sólo si sacamos primero una carta para ver si es rey y se vuelve a colocar en la baraja y luego sacamos una segunda carta para mirar si es oros.	64	32,7	61	14,7
c.) No, porque $P(\text{rey/oros}) \neq P(\text{rey})$.	17	8,7	82	19,8
d.) Sí, en todos los casos.	51	26,0	120	29,0
En blanco.	1	0,5	33	8,0
Total	196	100,0	414	100,0

Tabla 3 Respuestas al problema 3

En la Tabla 3 mostramos las respuestas de nuestros estudiantes a este problema, donde vemos que un 26% de futuros profesores y 29% de alumnos de Psicología responden correctamente. El error más frecuente entre nuestros alumnos fue confundir independencia con mutua exclusividad (alrededor del 30% en cada grupo), algo que coincide con los resultados de Sánchez ([6]). El distractor (b) evalúa el error de suponer que el concepto de independencia sólo se puede aplicar a experimentos que se suceden en el tiempo, fue cometido por un 14,7% de los estudiantes de psicología y en el doble de futuros profesores. Fue algo mayor la proporción de estudiantes de psicología que piensan que los sucesos no son independientes precisamente porque se cumple la relación de independencia y también que no responden que los futuros profesores.

1.3 Condicionamiento, Causación y Temporalidad

La causalidad es un concepto científico, filosófico y psicológico complejo, a pesar de que intuitivamente es comprendido y aceptado por la mayoría de personas ya que construimos nuestro conocimiento del mundo sobre la base de relaciones de causa y efecto entre diferentes sucesos. Desde el punto de vista probabilístico, si un suceso A es causa de otro suceso B , siempre que suceda A , sucederá B , por lo que $P(B/A) = 1$. La relación causal estricta es difícil de hallar en el mundo real y hablamos de relación de *causa débil* cuando al suceder A cambia la probabilidad de que ocurra B . Es decir, cuando $P(B/A)$ es diferente de $P(B)$, por lo cual una relación de causalidad implica una dependencia de tipo estadístico entre los sucesos implicados.

Sin embargo, en contra de la creencia popular, dos sucesos pueden ser estadísticamente dependientes, sin que uno de ellos sea causa del otro. Por ejemplo, es sabido que los países con mayor esperanza de vida tienen una menor tasa de natalidad, pero esto no implica que la tasa de natalidad sea causa de la esperanza de vida o al contrario, ya que si conseguimos aumentar la natalidad de un país esto no incide automáticamente en la esperanza de vida de sus habitantes. La existencia de una relación condicional indica que una relación causal es posible, pero no segura. Una asociación estadística entre variables puede ser debida a otras variables intervinientes o incluso ser espúrea y no implica relación causal. En el ejemplo, ambas variables pueden depender del nivel de renta de un país que es mayor si hay mayor proporción de mujeres trabajando; este trabajo de las mujeres contribuye a elevar el nivel de renta y este la esperanza de vida. Pero la mayor dedicación de las mujeres a la vida profesional hace que estas decidan tener un menor número de hijos, disminuyendo la tasa de natalidad.

Desde el punto de vista psicológico, la persona que evalúa una probabilidad condicional $P(A/B)$ va a percibir dos relaciones muy diferentes entre A (suceso condicionado) y B (suceso condicionante) dependiendo del contexto. Si dentro del contexto se percibe que B es una causa de A , la persona establecerá entre A y B una *relación causal*, por ejemplo, al preguntar cuál es la probabilidad de que una niña tenga los ojos azules si su padre tiene los ojos azules. Si dentro del contexto se percibe A como una causa de B , la persona establecerá entre A y B una *relación diagnóstica*, por ejemplo, al preguntar cuál es la probabilidad de que un padre tenga los ojos azules si una niña tiene los ojos azules [7]. Aunque matemáticamente los dos enunciados son equivalentes, desde un punto de vista psicológico no son percibidos como idénticos por las personas. En la primera opción, A sería que la niña tenga los ojos azules (efecto), B que la madre tenga los ojos azules (causa). En este caso al calcular $P(A/B)$ el estudiante tendrá que realizar un razonamiento causal, estimando el efecto dado cierto conocimiento de las causas. Por el contrario, en la segunda opción, A sería que la madre tenga los ojos azules (causa) y B que la niña tenga los ojos azules, por tanto $P(A/B)$, es una relación diagnóstica.

Numerosos estudios indican que la creencia que las relaciones causales son más fuertes que las relaciones diagnósticas. La relación de causalidad también se asocia, a menudo, con la secuencia temporal. El problema 4 [8] ilustra como algunos estudiantes tienen problemas con la condicionalidad cuando se invierte el eje de tiempo en que los sucesos ocurren de una forma natural.

Problema 4. Una urna contiene dos bolas blancas y dos bolas negras. Extraemos a ciegas dos bolas de la urna, una detrás de otra, sin reemplazamiento.

1. ¿Cuál es la probabilidad de extraer una bola negra en segundo lugar, habiendo extraído una bola negra en primer lugar? $P(N2/N1)$
2. ¿Cuál es la probabilidad de extraer una bola negra en primer lugar, sabiendo que hemos extraído una bola negra en segundo lugar? $P(N1/N2)$

Mientras los alumnos de Falk [8] no tenían dificultad para resolver la primera parte del problema 4, muchos fueron incapaces de dar una solución a la segunda, a la que responden que el resultado en la segunda extracción no afecta a la primera. Otros estudiantes dan como respuesta $1/2$, teniendo en cuenta sólo la composición de la urna y sin utilizar el dato del resultado posterior. Estos resultados sugieren la confusión entre condicionamiento y causación. En la primera parte del problema 4, la inferencia causal es una situación natural y compatible con el eje temporal, pero la segunda parte pide hacer una inferencia inversa, que requiere un razonamiento probabilístico que es indiferente al orden temporal, lo que puede crear dificultades psicológicas, pues el resultado en la segunda extracción depende causalmente del resultado en la primera extracción, pero no al contrario. Sin embargo, el resultado en cualquiera de las dos extracciones modifica la estimación de probabilidades del resultado en la otra. Si, por ejemplo, en la

segunda extracción ha salido una bola negra, sabemos que esta bola ya no puede ser uno de los posibles resultados en la primera extracción, por tanto ha habido una reducción del espacio muestral y $P(N_1/N_2)$ sería un tercio, igual que la $P(N_2/N_1)$.

La creencia de que un suceso que ocurre después del que juzgamos no puede afectar a la probabilidad del primero se conoce como *falacia del eje temporal*. Es importante erradicar esta creencia, pues la probabilidad de un suceso debe ser revisada a la luz de resultados posteriores en algunas situaciones, sobre todo al aplicar el Teorema de Bayes, donde la actualización de las probabilidades a la luz de los resultados juega un papel tan importante [9].

	Futuros profesores (n = 196)		Psicología (n = 414)	
	Frecuencia	%	Frecuencia	%
a.) 1/ 2	7	3,6	23	5,6
b.) 1/ 6	52	26,5	70	16,9
c.) 1/ 3	126	64,3	285	68,8
d.) 1/ 4	8	4,1	30	7,3
En blanco.	3	1,5	6	1,4
Total	196	100,0	414	100,0

Tabla 4 Respuestas al problema 4 (parte 1)

En las tablas 4 y 5 presentamos los resultados de la aplicación del problema 4 en nuestros dos estudios. La mayoría de los alumnos de psicología (68,8%) y de futuros profesores (64,3 %) han respondido correctamente a la primera parte del problema. El error más frecuente en esta parte en ambos grupos fue confundir la probabilidad condicional y conjunta aplicando la regla del producto $(1/2) \times (1/3) = 1/6$.

	Futuros profesores (n = 196)		Psicología (n = 414)	
	Frecuencia	%	Frecuencia	%
a.) 1/3	58	29,6	99	23,9
b.) No se puede calcular	41	20,8	103	24,9
c.) 1/6	25	12,8	38	9,2
d.) 2	64	32,7	151	36,5
En blanco.	8	4,1	23	5,5
Total	196	100,0	414	100,0

Tabla 5 Respuestas al problema 4 (parte 2)

En la segunda parte del problema, son sólo el 29,6% de futuros profesores y el 23,9% de estudiantes de Psicología los que dan la respuesta correcta, mientras que un 20,8% y un 24,9% respectivamente indican que no se puede calcular, mostrando explícitamente la falacia del eje temporal. El sesgo de equiprobabilidad lo presenta un 32% y 36,5% de estudiantes y el resto confunde probabilidad condicional y conjunta (distractor c). Nuestros resultados reproducen los obtenidos en [8].

1.4 Intercambio de Sucesos en la Probabilidad Condicional

Falk [8] sugirió que muchos estudiantes no discriminan adecuadamente entre las dos direcciones de la probabilidad condicional $P(A/B)$ y $P(B/A)$ y denominó a este error *falacia de la condicional transpuesta*. Aparece con frecuencia en contextos médicos, donde se confunde la probabilidad de tener una enfermedad cuando ha sido positivo el test de diagnóstico con la probabilidad de un resultado positivo en el test de diagnóstico, dado que se tiene la enfermedad [10]. La prevalencia de este error puede tener consecuencias importantes; por ejemplo la confusión entre la probabilidad de que un niño afectado con síndrome de Down dé una amniocentesis prenatal positiva, que es alta y el hecho de que, siendo la prueba positiva el niño realmente tenga síndrome de Down, que es mucho menor. Un problema de este tipo se presentó a nuestros estudiantes para evaluar la presencia de este sesgo presentamos (Problema 5). Algo más del 40% de futuros profesores y alrededor de tercio de estudiantes de Psicología escoge la respuesta correcta (alternativa b), siendo los resultados en nuestro caso algo mejores que en el estudio de Pollatsek y cols. [7]. Lo más frecuente en ambos grupos es considerar la misma confianza en ambas predicciones, lo que de acuerdo a Pollatsek et al., indica la falacia de la condicional transpuesta.

Problema 5. Un test diagnóstico de cáncer fue administrado a todos los residentes de una gran ciudad en la que hay pocos casos de cáncer. Un resultado positivo en el test es indicativo de cáncer y un resultado negativo es indicativo de ausencia de cáncer. ¿Qué te parece más probable?

1. Que una persona tenga cáncer si ha dado positivo en el test de diagnóstico.
2. Que un test de diagnóstico resulte positivo si la persona tiene cáncer.
3. Los dos sucesos tienen la misma probabilidad.

	Futuros profesores ($n = 196$)		Psicología ($n = 414$)	
	Frecuencia	%	Frecuencia	%
a.) Predecir que una persona tiene cáncer si ha dado positivo en el test de diagnóstico.	27	13,8	24	5,8
b.) Predecir un resultado positivo en el test de diagnóstico si la persona tiene cáncer.	81	41,3	133	32,1
c.) Tengo la misma confianza en ambas predicciones.	83	42,3	245	59,2
En blanco.	5	2,6	12	2,9
Total	196	100,0	414	100,0

Tabla 6 Respuestas de futuros profesores y alumnos de Psicología al problema 5

La falacia de la condicional transpuesta aparece con frecuencia al interpretar el nivel de significación α en los contrastes de hipótesis. El nivel de significación α se define como la probabilidad condicional de obtener un resultado R en la región de rechazo cuando la hipótesis nula H_0 es cierta, es decir $\alpha = P(R/H_0)$. Cuando un contraste de hipótesis resulta significativo (lo que quiere decir que R ha ocurrido) y alguien pregunta por la probabilidad de haber cometido un error (la probabilidad de que H_0 sea cierta) a menudo se contesta con α . En esta situación se estaría confundiendo $P(R/H_0)$ con $P(H_0/R)$. También se ha encontrado en la interpretación de tablas de contingencia por parte de estudiantes donde, según [11] alrededor del 20% de los estudiantes del curso preuniversitario en su trabajo

confunden “porcentaje de fumadores que contraen cáncer de pulmón” con “porcentaje de personas con cáncer de pulmón que fuman”.

Una posible explicación dada por Falk [8] de la prevalencia de este error es que el lenguaje ordinario, que es el que usamos en el enunciado de los problemas de probabilidad condicional no tiene la suficiente precisión. Cuando escribimos una probabilidad condicional usando la notación matemática es claro cuál es el suceso condicionante y cuál el condicionado, pero en el lenguaje ordinario la probabilidad condicional (tener cáncer si se es fumador) y su inversa (ser fumador si se tiene cáncer) no siempre se distinguen claramente entre sí. También en el caso del contraste estadístico de hipótesis la definición de α como “probabilidad de cometer error tipo I” podría contribuir a su incorrecta interpretación porque en la anterior frase sólo se hace referencia a un suceso (error tipo I) y no a una probabilidad condicional.

1.5 Confusión de Probabilidad Condicional y Probabilidad Conjunta

Pollatsek et al. [7] coinciden con Falk [8] en que muchas de las dificultades que las personas tienen con la comprensión de la probabilidad condicional pueden deberse a dificultades de comprensión del lenguaje de los enunciados de la probabilidad condicional y que la ejecución de tareas que implican probabilidades condicionales depende de cómo se redacten los enunciados. Un ejemplo lo encontraron Einhorn y Hogarth [12] con los enunciados que usan la conjunción “y”. Para estos autores estos enunciados pueden llevar a confundir la probabilidad conjunta y la probabilidad condicional. Los autores plantearon a 24 estudiantes la pregunta: “¿Cuál es la probabilidad de ir al supermercado y comprar café?”. Esta pregunta se refiere a la probabilidad conjunta, pero 9 de los 24 estudiantes la interpretaron en forma condicional como $P(\text{comprar café} / \text{ir al supermercado})$. También en [13] la mitad de los sujetos del estudio interpretaron la intersección como condicionamiento.

Un error relacionado es la *falacia de la conjunción* [14] o creencia de que es más probable la intersección de dos sucesos que la de uno de ellos por separado o la de su unión. Según Tversky y Kahneman el error es resultado de considerar la conjunción como más representativa de la población generadora que cada evento separado o bien del hecho que la conjunción hace que los sujetos recuerden o imaginen más ejemplos de una categoría o modelo más restringido. Díaz [15], en su estudio sobre la falacia de la conjunción, hace hincapié en la influencia de los sucesos que intervienen en los problemas. Cuando uno de los sucesos tiene una probabilidad muy alta en comparación con el otro, la intersección de los dos se ve más probable que el suceso de mayor probabilidad, por lo que aparece la falacia. Un ejemplo sería el Problema 6, cuya solución correcta es la (a), ya que al ser sucesos independientes esta va a ser menor que la probabilidad dada en el distractor (b), en la que intervienen los sucesos ganar el segundo y tercer set.

Problema 6. Supón que Rafa Nadal alcanza la final de Roland Garros en 2004. ¿Cuál de los siguientes sucesos consideras más probable?

1. Rafa Nadal pierde el primer set
2. Rafa Nadal pierde el primer set pero gana el partido
3. Los dos sucesos son iguales de probables

	Futuros profesores ($n = 196$)		Psicología ($n = 414$)	
	Frecuencia	%	Frecuencia	%

a.) Rafa Nadal pierde el primer set.	112	57,1	103	24,9
b.) Rafa Nadal pierde el primer set pero gana el partido.	23	11,8	39	9,4
c.) Los dos sucesos son iguales de probables.	58	29,6	256	61,8
En blanco.	3	1,5	16	3,9
Total	196	100,0	414	100,0

Tabla 7. Respuestas al problema 6

Como vemos en la Tabla 7, el 57,1% de los futuros profesores dio una solución correcta; por tanto estos alumnos distinguen correctamente entre probabilidad conjunta, condicional y simple. Los resultados son bastante mejores en este ítem que los de Díaz, cuyo porcentaje de respuestas correctas fue sólo del 24,9%. Un 11,7% de futuros profesores y un 9,4% de estudiantes de Psicología optaron por la opción (b); se tratan de estudiantes que están incurriendo en la falacia de la conjunción. En consecuencia, la falacia de la conjunción tiene poca incidencia en ambos estudios. Un 29,6% de futuros profesores y un 61,8% de la de Psicología eligieron la opción (c), manifestando el sesgo de equiprobabilidad descrito en los experimentos de Lecoutre [16]. Esta autora describe la creencia de los sujetos en la equiprobabilidad de todos los sucesos asociados a cualquier experimento aleatorio, incluso en aquellos en que no es aplicable el principio de indiferencia o donde no hay una simetría física.

1.6 Situaciones Sincrónicas y Diacrónicas

Otra variable que influye en la dificultad de las tareas de probabilidad condicional es si se percibe o no el experimento compuesto como una serie de experimentos simples sucesivos. A este respecto se pueden distinguir dos tipos de situaciones relacionadas con la probabilidad condicional:

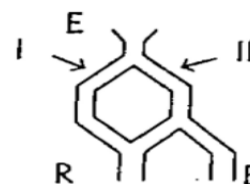
- *Situaciones sincrónicas:* Se trata de situaciones estáticas, en las que no subyace una secuencia de experimentos, sino que los experimentos aleatorios se realizan simultáneamente. Un ejemplo lo encontramos en el problema 2, en que se extrae una carta de una baraja pero se consideran dos experimentos (ver el palo y la figura de la carta).
- *Situaciones diacrónicas:* Son situaciones en las que hay una clara secuencia temporal, donde se realizan un experimento detrás de otro. Un ejemplo se muestra en el problema 1 en que se lanza varias veces seguidas un dado.

Cuando calculamos una probabilidad condicional, es clave que cambiemos el espacio muestral al suceso condicionante, pero algunos estudiantes encuentran difícil identificar el espacio muestral cuando hay un suceso condicionante. Las situaciones sincrónicas dificultan especialmente el cambio de espacio muestral, como comprueba Ojeda [13] al plantear un problema de este tipo a 26 alumnos de entre 14 y 16 años, el 60% de los cuáles hicieron una restricción incorrecta del espacio muestral.

La autora propone también los problemas 4a y 7 a 255 alumnos de Bachillerato, después de estudiar probabilidad condicional. Estos dos problemas tienen la misma estructura matemática, pero el problema 4a se percibe más fácilmente como una secuencia de experimentos. La proporción de respuestas correctas subió del 25% (problema 7) al 40% (problema 4a), aunque los alumnos todavía siguen con dificultades y no reducen convenientemente el espacio muestral al resolver el problema. En nuestro estudio también observamos un alto porcentaje de respuestas correctas tanto en futuros profesores como en estudiantes de Psicología en el ítem 7a.

Problema 7. Una bola se suelta en E. Si sale por R, ¿Cuál es la probabilidad de que haya pasado por el canal I?

1. 0,50
2. 0,33
3. **0,66**
4. No se puede calcular



En la tabla 8 presentamos los resultados del problema 7 en nuestros dos estudios. El resultado (c) es el correcto ya que para resolver este ítem, el alumno ha de tener en cuenta que si la bola pasa por el canal II puede caer por R o por B, y si pasa por I, sólo puede caer por R. El espacio muestral de este experimento sería: {(I,R), (II,R), (II,B)}. Puesto que a R llega el doble de bolas desde I que desde II, $P(I/R) = 2/3$. Esta solución es sin embargo obtenida por muy pocos estudiantes de Psicología y sólo la cuarta parte de los futuros profesores. Contrastan estos resultados con lo de Ojeda [13] quien obtiene mucho mejores resultados con este ítem que con los anteriores. La diferente forma en que se da el enunciado podría explicar la discrepancia de resultados.

	Futuros profesores (n = 196)		Psicología (n = 414)	
	Frecuencia	%	Frecuencia	%
a.) 0,50	132	67,3	318	76,8
b.) 0,33	15	7,7	37	8,9
c.) 0,66	46	23,5	41	9,9
d.) No se puede calcular.	2	1,0	7	1,7
En blanco.	1	0,5	11	2,7
Total	196	100,0	414	100,0

Tabla 8 Respuestas al problema 7

La mayoría de los estudiantes en ambos estudios elige el distractor (a) y por tanto, no está teniendo en cuenta las bolas que caen por el orificio B, es decir no se tiene en cuenta el condicionamiento por un suceso posterior, haciendo una incorrecta restricción del espacio muestral.

El resto de los estudiantes elige el distractor (b) se está confundiendo el suceso condicionado ya que se está calculando la probabilidad de que habiendo salido la bola por R, la bola haya pasado por II o bien indica que no se puede calcular la probabilidad.

1.7 Implicaciones Para la Enseñanza de la Estadística

En este trabajo hemos descrito algunos sesgos que aparecen en las investigaciones en Psicología y Educación sobre comprensión de la probabilidad condicional y mostramos nuestros propios datos en dos estudios empíricos que muestran la extensión de estos sesgos no sólo en estudiantes de Psicología, sino también en futuros profesores de matemáticas.

Como vemos, el porcentaje que incurre en los diferentes sesgos es alto, tanto en el estudio de Díaz [1] como en el de Contreras [2]. Es especialmente preocupante la presencia de estos sesgos en los futuros profesores. Por ejemplo,

hay una gran incidencia de la falacia del eje de tiempos por la que los sujetos tienen dificultad para resolver un problema de probabilidad condicional si el suceso condicionante ocurre después del condicionado. Sin embargo, esta es una situación muy común en la vida real e incluso en el trabajo profesional del futuro profesor (por ejemplo en la evaluación de los alumnos, el suceso condicionante sería el resultado del examen y el condicionado los conocimientos del estudiante). También aparece con frecuencia en inferencia estadística donde conceptos como intervalo de confianza o nivel de significación están definidos mediante condicionales en que el eje de tiempos se ha invertido.

También observamos dificultad en la percepción de los experimentos compuestos en el caso de situaciones sincrónicas. Se confunde independencia y exclusión, se cambian los términos de la probabilidad condicional, se confunde ésta con la conjunta y se asigna a la probabilidad conjunta un valor mayor que a la probabilidad simple, violando las reglas lógicas del cálculo de probabilidades.

Estos resultados son motivo de preocupación, ya que los futuros profesores de nuestra muestra tenderán a fallar en la enseñanza de la probabilidad así como en algunas actividades profesionales que requieren del razonamiento probabilístico, tales como "*averiguar lo que los estudiantes saben, la elección y la gestión de las representaciones de las ideas matemáticas, la selección y modificación de los libros de texto, decidir entre los cursos alternativos de acción*" [17, p. 453]. En consecuencia, se sugiere la necesidad de mejorar la educación sobre probabilidad que estos futuros profesores reciben durante su formación.

Como indicamos en la introducción, la probabilidad condicional es un concepto básico para la comprensión de muchos otros que tiene que enseñar el futuro profesor de educación secundaria y Bachillerato. Las distribuciones de los estadísticos en el muestreo, el nivel de significación y potencia en un contraste de hipótesis, las distribuciones marginales, las rectas de regresión, entre otros conceptos, se definen mediante una probabilidad condicional. Es, por tanto, necesario asegurarnos que el profesor supera sus sesgos sobre este tema, si queremos evitar que los transmita a sus estudiantes.

Por otro lado, la compleja relación entre los conceptos probabilísticos y la intuición [18] se muestran también en los resultados, puesto que la alta preparación matemática no fue suficiente para evitar sesgos de razonamiento. Al igual que en otras investigaciones [19] los resultados sugieren la necesidad de prestar más atención a la enseñanza de heurísticas en la resolución de problemas matemáticos y la importancia que en la resolución de los problemas matemáticos tienen los procesos psicológicos.

Los problemas que hemos mostrado como ejemplos a lo largo del trabajo pueden usarse para diagnosticar las dificultades de los futuros profesores y organizar acciones formativas que les ayuden a superarlos. Las soluciones erróneas pueden discutirse colectivamente y la simulación de las experiencias descritas en los problemas, con ayuda de tablas de números aleatorios, calculadoras u ordenadores permitirá la superación gradual de estos sesgos. El uso de diversas representaciones, como árboles, o diagramas rectangulares puede también contribuir a la mejora del aprendizaje.

Agradecimientos: Este trabajo forma parte del proyecto: EDU2010-14947, MICINN- FEDER y grupo PAI FQM126.

Bibliografía

- [1] Díaz, C. (2007). *Viabilidad de la inferencia bayesiana en el análisis de datos en psicología*. Universidad de Granada: Tesis doctoral.

- [2] Contreras, J. M. (2011). *Evaluación de conocimientos y recursos didácticos en la formación de profesores sobre probabilidad condicional*. Tesis Doctoral. Universidad de Granada.
- [3] Maury, S. (1985). Influence de la question dans una épreuve relative á la notion d'independance. *Educational Studies in Mathematics*, 16, 283-301.
- [4] Maury, S. (1986). *Contribution à l'étude didactique de quelques notions de probabilité et de combinatoire à travers la résolution de problèmes*. Tesis doctoral. Universidad de Montpellier II.
- [5] Kelly, I. W. y Zwiers, F. W. (1986). Mutually exclusive and independence: Unravelling basic misconceptions in probability theory. *Teaching Statistics*, 8, 96-100.
- [6] Sánchez, E. (1996). Dificultades en la comprensión del concepto de eventos independientes. En F. Hitt (Ed.), *Investigaciones en Educación Matemática* (pp. 389-404). México.
- [7] Pollatsek, A., Well, A. D., Konold, C. y Hardiman, P. (1987). Understanding Conditional Probabilities. *Organization, Behavior and Human Decision Processes*. 40, 255 – 269.
- [8] Falk, R. (1986). Conditional probabilities: insights and difficulties. En R. Davidson y J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics*. (pp. 292–297). Victoria, Canada: International Statistical Institute.
- [9] Gras, R. y Totohasina, A. (1995). Chronologie et causalité, conceptions sources d'obstacles épistémologiques à la notion de probabilité conditionnelle *Recherches en Didactique des Mathématiques*, 15(1), 49-95.
- [10] Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. En D. Kahneman, P. Slovic y Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases*. New York: Cambridge University Press.
- [11] Batanero, C., Estepa, A., Godino, J. y Green, D. R. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education*, 27(2), 151–169.
- [12] Einhorn, H. J. y Hogart, R.M. (1986). Judging probable cause. *Psychological Bulletin*, 99, 3–19.
- [13] Ojeda, A. M. (1995). Dificultades del alumnado respecto a la probabilidad condicional. *UNO*, 5, 37-55.
- [14] Tversky, A. y Kahneman, D. (1982). Causal schemas in judgment under uncertainty. En D. Kahneman, P. Slovic y A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases* (pp. 117-128). Cambridge, MA: Cambridge University Press.
- [15] Díaz, C. (2005). Evaluación de la falacia de la conjunción en alumnos universitarios. *Suma*, 48, 45-50.
- [16] Lecoutre, M. P. (1992). Cognitive models and problem spaces in purely random situations. *Educational Studies in Mathematics*, 23, 557-568.
- [17] Ball, D. L., Lubienski, S. T., y Mewborn, D. S. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. En V. Richardson (Ed.), *Handbook of research on teaching* (pp. 433-456). Washington, DC: American Educational Research Association.
- [18] Borovcnik, M. y Peard, R. (1996). Probability. En A. Bishop, et al. (Eds.), *International handbook of mathematics education* (pp. 239-288). Dordrecht: Kluwer.
- [19] Aguilar, M., Navarro, J.I., López, J.M. y Alcalde, C. (2002). Pensamiento formal y resolución de problemas matemáticos. *Psicothema*, 14(2), 382-386.