Modelos Probabilísticos

Ayudantía 2

Eduardo Alarcón Bustamante

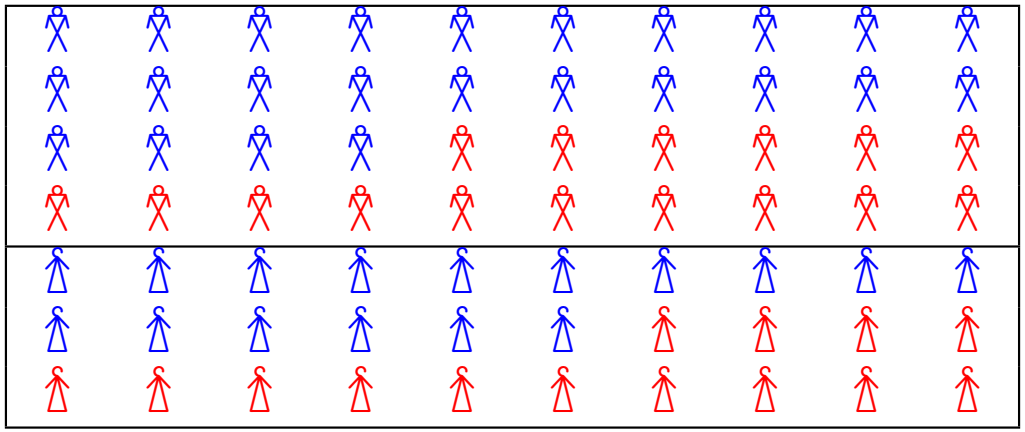
Departamento de Estadística,
Pontificia Universidad Católica de Chile.

Segundo semestre - 2022

Herramientas básicas del modelamiento estructural

- El concepto de probabilidad condicional, el teorema de probabilidades totales y el teorema de Bayes-Laplace, son las herramientas básicas que permiten especificar lo que en Estadística llamamos *modelos estructurales*.
- A continuación introducimos estas herramientas usando ejemplos sencillos, enfatizando los aspectos semánticos.

Consideremos 70 estudiantes (40 varones y 30 mujeres) Para cada estudiante, se observa si fue admitido en una determinada universidad (en azul) o no (en rojo).



Probabilidades marginales y condicionales

¿Cuál es la proporción de estudiantes que fue admitido en la universidad?

$$P(\text{estudiante admitido}) = \frac{40}{70}$$

¿Cuál es la proporción de estudiantes que no fue admitido en la universidad?

$$P(\text{estudiante no admitido}) = \frac{30}{70}$$

Las probabilidades anteriores se calculan teniendo en cuenta toda la población de postulantes: **por eso se llaman probabilidades marginales**.

Las proporciones de estudiantes admitidos y no admitidos pueden reportarse también bajo una determinada condición, por ejemplo el sexo biológico.

¿Cuál es la proporción de mujeres que fue admitido en la universidad?

$$P(\text{estudiante admitido} \mid \text{♀}) = \frac{24}{40}$$

¿Cuál es la proporción de hombres que fue admitido en la universidad?

$$P(\text{estudiante admitido} \mid \text{♂}) = \frac{16}{30}$$

Cada una de estas probabilidades se calcula con respecto a la **población definida por la condición**, por lo que se llaman **probabilidades condicionales**.

Ley de probabilidades totales

- Es posible relacionar la **probabilidad marginal** $P(\text{Estudiante admitido})$ con las **prob condicionales** $P(\text{estudiante admitido} \mid \text{🧍‍♀️})$ y $P(\text{estudiante admitido} \mid \text{🧍})$.
- Para ello solo es necesario considerar la proporción de hombres y de mujeres entre los encuestados:

$$P(\text{🧍}) = \frac{30}{70} \quad P(\text{🧍‍♀️}) = \frac{40}{70}$$

$$
\begin{aligned}
P(\text{Est admitido}) &= \frac{40}{70} \\
&= \frac{24+16}{70} \quad \text{pues hay 24 varones y 16 mujeres que fueron admitidos} \\
&= \frac{24}{70} + \frac{16}{70} \\
&= \frac{24}{40} \times \frac{40}{70} + \frac{16}{30} \times \frac{30}{70} \quad \text{pues hay 40 varones y 30 mujeres} \\
&= P(\text{Est admitido} \mid \text{🚹}) \, P(\text{🚹}) + P(\text{Est admitido} \mid \text{🚺}) \, P(\text{🚺}).
\end{aligned}
$$

La descomposición anterior siempre se puede hacer: basta descomponer el espacio muestral en *subpoblaciones o grupos*, y luego caracterizar las probabilidades condicionales de una variable de interés para cada grupo. Esto se llama **Ley de Probabilidades Totales** y no es nada más que un *promedio ponderado de probabilidades condicionales*.

Descomposición marginal-condicional

- Antes de introducir el Teorema deBayes-Laplace, vamos a formalizar un concepto muy importante en el modelamiento estructural, a saber *descomposición marginal condicional*.
- A fin de hacer explícita dicha descomposición, usaremos un argumento debido a Laplace (1820, pp. 181-182).

Supongamos que el número total de casos posibles que pueden ocurrir es igual a p. Supongamos además que

1. De los p casos posibles, hay p' casos favorables a la ocurrencia de un evento E_1.
2. De estos p' casos, hay q casos favorables a la ocurrencia del evento E_2.

A modo de ejemplo, consideremos que:

- E_1= "día nublado en un año determinado"
- E_2 = "día lluvioso en un año determinado"

Así, si llueve, entonces necesariamente está nublado (el "estar nublado" queda determinado por el día lluvioso), pero si está nublado no necesariamente llueve: llover es un evento contingente a estar nublado.

Notemos que

- $P(E_1) = \dfrac{p'}{p}$: proporción de días nublados en un año determinado.
- $P(E_2 \mid E_1) = \dfrac{q}{p'}$: proporción de días lluviosos entre los días nublados.
- $P(E_1 \cap E_2) = \dfrac{q}{p}$: proporción de días lluviosos (y por tanto nublados) en un año determinado.

Note que

$$P(E_1 \cap E_2) = \frac{p'}{p} \times \frac{q}{p'} = P(E_1)P(E_2 \mid E_1)$$

Esto se conoce como una descomposición *marginal-condicional*: indica en qué medida las condiciones definidas por el evento E_1 determinan la probabilidad de ocurrencia del evento E_2.

Teorema de Bayes-Laplace

Retomemos el ejemplo utilizado y supongamos que extraemos al azar un estudiante y resulta que fue admitido en la universidad. ¿Cuál es la probabilidad que dicho estudiante sea una mujer?
Información disponible:

$$P(\text{estudiante admitido}) = \frac{40}{70}; \quad P(\text{estudiante admitido} \mid \text{🧍}) = \frac{24}{40}; \quad P(\text{🧍}) = \frac{30}{70}$$

Sin embargo, necesitamos

$$P(\text{🧍} \mid \text{estudiante admitido})$$

Notemos que la descomposición marginal condicional, nos permite escribir

$$P(\text{🧍} \cap \text{estudiante admitido}) = P(\text{🧍} \mid \text{estudiante admitido})P(\text{estudiante admitido})$$
$$P(\text{🧍} \cap \text{estudiante admitido}) = P(\text{estudiante admitido} \mid \text{🧍})P(\text{🧍})$$

Entonces

$$P(\text{🧍} \mid \text{estudiante admitido})P(\text{estudiante admitido}) = P(\text{estudiante admitido} \mid \text{🧍})P(\text{🧍})$$
$$P(\text{🧍} \mid \text{estudiante admitido}) = \frac{P(\text{estudiante admitido} \mid \text{🧍})P(\text{🧍})}{P(\text{estudiante admitido})}$$
$$= \frac{24/40 \times 30/70}{40/70} = 0.45 \qquad (1)$$

Notemos que por el teorema de probabilidad total,

$$P(\text{Est admitido}) = P(\text{Est admitido} \mid \text{🧍‍♀️})\, P(\text{🧍‍♀️}) + P(\text{Est admitido} \mid \text{🧍‍♂️})\, P(\text{🧍‍♂️}).$$

Entonces,

$$P(\text{🧍‍♂️} \mid \text{Est admitido}) \quad = \quad \frac{P(\text{Est admitido} \mid \text{🧍‍♂️}) P(\text{🧍‍♂️})}{P(\text{Est admitido} \mid \text{🧍‍♀️})\, P(\text{🧍‍♀️}) + P(\text{Est admitido} \mid \text{🧍‍♂️})\, P(\text{🧍‍♂️})}$$

La relación anterior se conoce como el Teorema de Bayes-Laplace.

En términos más generales, el Teorema de Bayes-Laplace se enuncia de la siguiente manera. Sea E un evento que depende de n eventos mutuamente excluyentes entre sí $C_1, C_2, \ldots, C_n$. Por el Teorema de Probabilidades Totales sabemos que

$$P(E) = P(E \mid C_1)P(C_1)P(E \mid C_2)P(C_2) + \cdots P(E \mid C_n)P(C_n).$$

Esta descomposición significa que la ocurrencia del evento E depende de los n eventos y por tanto su probabilidad de ocurrencia $P(E)$ es una suerte de promedio ponderado de las probabilidades $P(E \mid C_i)$, donde los ponderadores corresponden a las probabilidades $P(C_i)$'s.

Ahora bien, si suponemos que el evento E se ha realizado, entonces nos preguntamos cuál es la probabilidad de que haya sido el evento C_i el que lo produjo. El Teorema de Bayes permite calcular esta probabilidad:

$$P(C_i \mid E) = \frac{P(E \mid C_i)P(C_i)}{P(E \mid C_1)P(C_1) + P(E \mid C_2)P(C_2) + \cdots P(E \mid C_n)P(C_n)}.$$

Aplicación: Test diagnósticos

Las probabilidades condicionales son el leguaje adecuado para estudiar las relaciones que hay entre el diagnóstico proprocionado por un test y el estado real de los dignosticados. Cuando una persona es diagnosticada con un test, no necesariamente el test lo diagnostica correctamente. Como ejemplo, consideremos la relación entre los resultados de un test (escáner de hígado) y el diagnóstico correcto basado en una biopsia. Los resultados se reumen en la siguiente tabla:

Escáner	Patología anormal $P = +$	Patología Normal $P = -$
$S = +$	231	32
$S = -$	27	54

¿Cuán bueno es el escáner de hígado para diagnosticar la patología? Para responder a esta pregunta, podemos calcular la proporción de pacientes con escáner de hígados normales y anormales que son correctamente diagnosticados por el escáner. Así, hay 258 pacientes que efectivamente tienen una patología anormal (enfermos del hígado) y 86 que efectivamente tienen una patología normal (sanos del hígado), a los que se les designa *verdaderos positivos* y *verdaderos negativos*, respectivamente. Las proporciones de estos dos grupos que fueron correctamente diagnosticados por el escáner son

$$\frac{231}{258} = 0.90, \qquad \frac{54}{86} = 0.63,$$

respectivamente. Estas dos proporciones son dnominadas como:

1. *Sensitividad* es la proporción de verdaderos positivos que son correctamente identificados por el test.
2. *Especificidad* es la proporción de verdaderos negativos que son correctamente identificados por el test.

En términos de probabilidades condicionales, se tiene lo siguiente:

1. Para la sensitividad, la condición bajo la cual se calcula la proporción de personas correctamente identificadas por el test es *personas que son verdaderos positivos* $(P = +)$. Por lo tanto,

$$\text{Sensitividad} \doteq P(S = + \mid P = +).$$

2. Para la especificidad, la condición bajo la cual se calcula la proporción de personas correctamente identificadas por el test es *personas que son verdaderos negativos* $(P = -)$. Por lo tanto,

$$\text{Especificidad} \doteq P(S = - \mid P = -).$$

La sensitividad y especificidad constituyen una aproximación para cuantificar la capacidad de un test diagnóstico. En la práctica clínica, sin embargo, solo se conocen los resultados del test, por lo que interesa conocer cuán bueno es el test prediciendo la enfermedad.

Esto nos lleva a introducir las siguientes definiciones:

1. *Valor predictivo positivo* es la proporción de pacientes con resultados positivos en el test que son correctamente diagnosticados; esto es,

$$\text{Valor predictivo positivo } \doteq P(P = + \mid S = +).$$

Usando los datos, se tiene que

$$P(P = + \mid S = +) = \frac{231}{263} = 0.88.$$

2. *Valor predictivo negativo* es la proporción de pacientes con resultados negativos en el test que son correctamente diagnosticados; esto es,

$$\text{Valor predictivo negativo } \doteq P(P = - \mid S = -).$$

Usando los datos, se tiene que

$$P(P = - \mid S = -) = \frac{54}{81} = 0.59.$$

Aplicación: Sesgo de selección

Ejemplo: a finales de Marzo de 2020, se publicó la encuesta Cadem, la cual reporta que el 21% **de los encuestados** aprueba la gestión del Presidente de la República. La encuesta proporciona la siguiente información adicional: se contactaron 2818 personas, de las cuales solo 707 respondieron. Por tanto, ¿qué información se provee cuando se dice que *el 21% de los encuestados aprueba la gestión del Presidente*?

Sea A el evento *Aprueba gestión del presidente* y sea R el evento *Responde la encuesta*, entonces:

$$P(A \mid R) = 0.21; \quad P(R) = \frac{707}{2818} = 0.25$$

Sin embargo, generalmente escuchamos que basado en la encuesta, se puede decir que el 21% de **los chilenos y chilenas** *Aprueba gestión del presidente*, esto es:

$$P(A) = P(A \mid R)$$

Por el teorema de probabilidades totales, tenemos que

$$P(A) = P(A \mid R)P(R) + P(A \mid R^c)P(R^c)$$

Para CADEM: *Bajo el supuesto de que* **quienes rechazan contestar son iguales a quienes contestan**,

$$P(A \mid R) = P(A \mid R^c)$$

la magnitud de la tasa de rechazo no ofrece mayores inconvenientes, pero cuando existe evidencia que ambos grupos no son equivalentes, el rechazo puede introducir serias distorsiones en los resultados.

¿Podemos comprobar este supuesto?

Si le creemos a este supuesto, entonces

$$P(A) = P(A \mid R)$$

Así, podemos obtener conclusiones de la población de interés (todos los chilenos y chilenas), solo con quienes accedieron a contestar (Diseño metodológico CADEM).

¿Y si no le creemos a este supuesto? Por los axiomas de Kolmogorov, lo único que sabemos es que $0 \leq P(A \mid R^c) \leq 1$

- $P(A \mid R^c) = 0$ Entre quienes no accedieron a contestar, NADIE aprueba la gestión del presidente, entonces

$$P(A) = P(A \mid R)P(R) = 0.21 \times 0.25 = 0.0525$$

- $P(A \mid R^c) = 1$ Entre quienes no accedieron a contestar, TODOS aprueba la gestión del presidente, entonces

$$P(A) = P(A \mid R)P(R) + P(R^c) = 0.21 \times 0.25 + 0.75 = 0.8025$$

Por lo tanto

$$0.0525 \leq P(A) \leq 0.8025$$

Por lo tanto, según lo entregado por la encuesta CADEM a finales de marzo, lo único que podemos decir es que *al menos un 5% de los chilenos y chilenas aprueba la gestión del Presidente, y a lo más un 80%*.

Veamos otro ejemplo acá

Paradoja de Simpson

A 40 pacientes se les administró un tratamiento T, y 40 pacientes fueron asignados al grupo de control, denotado por $\neg T$. Los pacientes se recuperan, R, o no, $\neg R$:

Table 1: Recuperaciones bajo tratamiento y control

	R	$\neg R$	Total
T	20	20	40
$\neg T$	16	24	40
Total	36	44	80

$$P(R \mid T) = \frac{20}{40} = 0.5; \quad P(R \mid \neg T) = \frac{16}{40} = 0.4.$$

Es decir,

$$P(R \mid T) > P(R \mid \neg T),$$

lo que induce a concluir que el tratamiento es preferible.

Sin embargo, el género de los pacientes también fue recolectado:

Table 2: Recuperaciones bajo tratamiento y control para varones

Varones	R	$\neg R$	Total
T	18	12	30
$\neg T$	7	3	10
Total	25	15	40

Table 3: Recuperaciones bajo tratamiento y control para mujeres

Mujeres	R	$\neg R$	Total
T	2	8	10
$\neg T$	9	21	30
Total	11	29	40

Entonces:

$$P(R \mid T, V) = \frac{18}{30} = 0.6, \quad P(R \mid \neg T, V) = \frac{7}{10} = 0.7,$$

$$P(R \mid T, M) = \frac{2}{10} = 0.2, \quad P(R \mid \neg T, M) = \frac{9}{30} = 0.3.$$

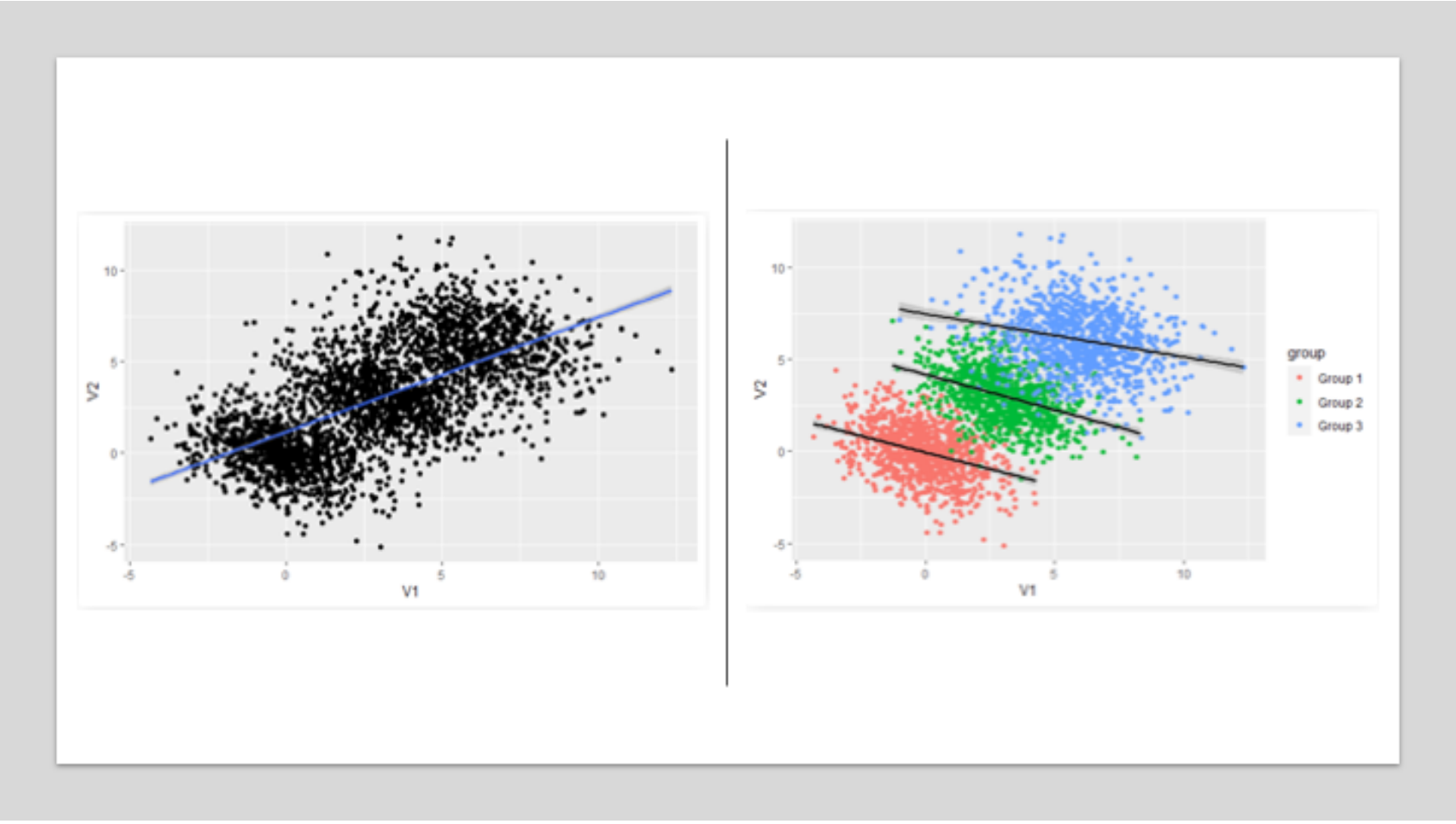
Es decir,

$$P(R \mid \neg T, V) > P(R \mid T, V) \quad \text{y} \quad P(R \mid \neg T, M) > P(R \mid T, M);$$

esto es, para cada grupo, varones y mujeres, no se prefiere el tratamiento: hay mayor probabilidad de recuperarse no recibiendo el tratamiento.

Este fenómeno es conocido como la *Paradoja de Simpson*.

La paradoja de Simpson en predicciones



Eventos independientes

Sean $A, B \subset \Omega$ dos eventos. La descomposición marginal-condicional implica que

$$P(A \cap B) = P(A)\, P(B \mid A) = P(B)\, P(A \mid B).$$

Esta descomposicón significa que A y B son eventos dependientes en el sentido que *la ocurrencia de uno de ellos determina la ocurrencia del otro*.

Por lo tanto, si el evento B no afectan la realización del evento A, entonces $P(A \mid B) = P(A)$. Cuando esto ocurre, se dice que los eventos A y B son *independientes*, lo que denotamos como $A \perp\!\!\!\perp B$.

Si $A \perp\!\!\!\perp B$, con $P(B) > 0$, entonces

$$P(A) = P(A \mid B) = \frac{P(A \cap B)}{P(B)} \Leftrightarrow P(A \cap B) = P(A)\, P(B).$$

Por lo tanto,

$$\begin{aligned} A \perp\!\!\!\perp B &\iff P(A \mid B) = P(A) \\ &\iff P(A \cap B) = P(A)\, P(B) \\ &\iff P(B \cap A) = P(B)\, P(A) \\ &\iff B \perp\!\!\!\perp A. \end{aligned}$$

Por lo tanto, la independencia entre eventos es simétrica, lo que permite escribir $P(A \cap B) = P(A \mid B)\, P(B \mid A)$. La independencia entre eventos satisface la siguiente invarianza:

Sean $A, B \subset \Omega$. Entonces

$$A \perp\!\!\!\perp B \iff A^c \perp\!\!\!\perp B \iff A \perp\!\!\!\perp B^c \iff A^c \perp\!\!\!\perp B^c.$$

Ejercicio: Demuestre esta propiedad.

Ejercicios sobre técnicas de conteo

Ej.1 Suponga que lanza un dado 7 veces y desea saber la probabilidad de obtener 3 veces el número cinco en estos 7 lanzamientos.

Definamos A: Obtener un 5.

Notemos que obtener 3 veces el número cinco significa lo siguiente

Obt 3 cincos $\Rightarrow \{A \cap A \cap A \cap A^c \cap A^c \cap A^c \cap A^c\}$
Obt 3 cincos $\Rightarrow \{A \cap A \cap A^c \cap A^c \cap A^c \cap A^c \cap A\}$
Obt 3 cincos $\Rightarrow \{A \cap A^c \cap A^c \cap A^c \cap A^c \cap A \cap A\}$
$\vdots$
Obt 3 cincos $\Rightarrow \{A^c \cap A^c \cap A \cap A^c \cap A \cap A \cap A\}$

Observación: requiere obtener 3 *cinco* y 4 *distintos de cinco*. Pero no se considera el orden en que se obtienen estos.

Por la observación anterior se tienen $\binom{7}{3} = 35$ formas de obtener 3 veces el número cinco en 7 lanzamientos.

Así

$$P(\text{Obt 3 cincos}) = 35 \cdot P(A \cap A \cap A^c \cap A^c \cap A^c \cap A^c \cap A)$$

Supongamos ahora que los lanzamientos se producen de forma independiente (El resultado de uno, no afecta la probabilidad del resultado que viene), entonces

$$\begin{aligned} P(\text{Obt 3 cincos}) &= 35 \cdot P(A)^3 \cdot [P(A^c)]^4 \\ &= 35 \cdot P(A)^3 \cdot [1 - P(A)]^4 \\ &= 35 \cdot (1/6)^3 \cdot (5/6)^4 \\ &= 0.0781 \end{aligned}$$

Ej.2 Suponga ahora que desea calcular la probabilidad de que se hayan necesitado 7 lanzamientos para observar por tercera vez el número cinco.