

## 6. Experimentos Compuestos

February 21, 2019

### 6. Experimentos Compuestos. Pruebas de Bernoulli

#### Experimentos Compuestos

Con frecuencia estamos interesados en experimentos aleatorios que acontecen simultánea o sucesivamente. Tales experimentos se modelan con sus propios **espacios de probabilidad marginales**, pero al interesarnos la relación entre ellos es necesario también modelar su **espacio de probabilidad conjunto**.

En algunos casos, **no existe dependencia probabilística entre los experimentos marginales**. En tal caso, **puede obtenerse fácilmente la caracterización conjunta a partir de las marginales**. Es el caso, por ejemplo, en el que lanzamos simultáneamente varios dados, o en el que lanzamos sucesivamente varias veces la misma moneda.

En otros, **sí existe dependencia probabilística entre los experimentos marginales**, en cuyo caso **se necesita conocer la caracterización conjunta, que no es posible obtener a partir de las marginales**. Por ejemplo, un espacio de probabilidad puede modelar que una persona tenga una enfermedad y otro el resultado de un test diagnóstico.

Consideremos que realizamos los experimentos  $\epsilon_1$ : "lanzar una moneda" y  $\epsilon_2$ : "lanzar un dado". Tanto el dado como la moneda son buenos (no están trucados). Ambos experimentos inducen sendos espacios de probabilidad:

- $\epsilon_1(\Omega_1, \mathcal{F}_1, P_1)$ , donde:
  - $\Omega_1 = \{c, x\}$ ,
  - $\mathcal{F}_1$  es el conjunto con los  $2^2 = 4$  sucesos posibles, y
  - $P_1(c) = P_1(x) = 1/2$
- $\epsilon_2(\Omega_2, \mathcal{F}_2, P_2)$ , donde:
  - $\Omega_2 = \{1, 2, 3, 4, 5, 6\}$ ,
  - $\mathcal{F}_2$  es el conjunto con los  $2^6 = 64$  sucesos posibles, y
  - $P_2(1) = P_2(2) = \dots = P_2(6) = 1/6$

El espacio muestral compuesto se obtiene mediante el **producto cartesiano**  $\Omega_1 \times \Omega_2$  de los espacios muestrales marginales,  $\Omega_1$  y  $\Omega_2$ , esto es, el conjunto formado por todos los **pares ordenados** que podemos obtener a partir de los elementos de ambos:

$$\Omega_1 \times \Omega_2 = \{(c, 1), (c, 2), (c, 3), (c, 4), (c, 5), (c, 6), \\ (x, 1), (x, 2), (x, 3), (x, 4), (x, 5), (x, 6)\}$$

| $\Omega_1 \times \Omega_2$ | 1        | 2        | 3        | 4        | 5        | 6        |
|----------------------------|----------|----------|----------|----------|----------|----------|
| $c$                        | $(c, 1)$ | $(c, 2)$ | $(c, 3)$ | $(c, 4)$ | $(c, 5)$ | $(c, 6)$ |
| $x$                        | $(c, 1)$ | $(c, 2)$ | $(c, 3)$ | $(c, 4)$ | $(c, 5)$ | $(c, 6)$ |

Sobre el espacio muestral conjunto puede definirse un álgebra de sucesos, consistente en todos los subconjuntos que puedan formarse a partir del mismo (en general, con ciertas restricciones que ahora no nos preocupan). Este álgebra de sucesos puede obtenerse con el producto cartesiano de las álgebras de sucesos de los espacios de probabilidad marginales,  $\mathcal{F}_1 \times \mathcal{F}_2$ .

En nuestro caso, tenemos  $2^6 = 64$  sucesos en cada espacio muestral, por lo que habrá  $2^{12} = 1024$  sucesos en el espacio de probabilidad conjunto. Algunos sucesos relevantes son:

- Suceso imposible:  $\emptyset = \{(A_1, \emptyset), (\emptyset, A_2), (\emptyset, \emptyset)\}$
- Suceso seguro:  $\Omega_1 \times \Omega_2$
- Suceso marginal  $A_1$  en primer experimento:  $A_1 \times \Omega_2$
- Suceso marginal  $A_2$  en segundo experimento:  $\Omega_1 \times A_2$
- Suceso  $A_1 \times A_2 = (A_1 \times \Omega_2) \cap (\Omega_1 \times A_2)$

Seguidamente, asignamos probabilidades a los elementos del espacio muestral conjunto.

| $P$ | 1         | 2         | 3         | 4         | 5         | 6         |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|
| $c$ | $P(c, 1)$ | $P(c, 2)$ | $P(c, 3)$ | $P(c, 4)$ | $P(c, 5)$ | $P(c, 6)$ |
| $x$ | $P(c, 1)$ | $P(c, 2)$ | $P(c, 3)$ | $P(c, 4)$ | $P(c, 5)$ | $P(c, 6)$ |

Y, con ello, es posible asignar probabilidades a los sucesos de  $\mathcal{F}_1 \times \mathcal{F}_2$ :

- Suceso imposible:  $P(A_1, \emptyset) = P(\emptyset, A_2) = P(\emptyset, \emptyset) = 0$
- Suceso seguro:  $P(\Omega_1 \times \Omega_2) = 1$ . La suma de probabilidades es 1.
- Cálculos de probabilidades marginales a partir de probabilidades conjuntas:
  - $P_1(A_1) = P(A_1 \times \Omega_2)$
  - $P_2(A_2) = P(\Omega_1 \times A_2)$
  - $P(A_1 \times A_2) = P((A_1 \times \Omega_2) \cap (\Omega_1 \times A_2))$
- En general, no puede calcularse la probabilidad conjunta a partir de las probabilidades marginales.

Por tanto, es posible definir un experimento de probabilidad compuesto, cuyo espacio de probabilidad conjunto se modela a partir de los marginales:

$$\epsilon_1 \times \epsilon_2 = \{\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, P\}$$

Las probabilidades conjuntas, en general, no están determinadas por las marginales:

$$P(A_1 \times A_2) = P((A_1 \times \Omega_2) \cap (\Omega_1 \times A_2))$$

Sin embargo, considerando que dos experimentos son independientes si y sólo si lo son todos sus sucesos, **en caso de independencia es posible calcular las probabilidades conjuntas a partir de las marginales:**

$$P(A_1 \times A_2) = P((A_1 \times \Omega_2))P((\Omega_1 \times A_2)) = P_1(A_1)P_2(A_2)$$

En el caso de nuestro ejemplo de experimento compuesto, consideremos el suceso  $(c, 3)$ :

$$P(\{c\} \times \Omega_2) = P_1(c) = 1/2 \quad P(\Omega_1 \times \{3\}) = P_2(3) = 1/6$$

Y las probabilidades de elementos arbitrarios del espacio muestral:

$$P((c, 3)) = P((\{c\} \times \Omega_2) \cap (\Omega_1 \times \{3\}))$$

En este experimento compuesto **podemos asumir independencia**, por tanto:

$$P((c, 3)) = P(\{c\} \times \Omega_2)P(\Omega_1 \times \{3\}) = P_1(c)P_2(3) = \frac{1}{2} \frac{1}{6} = \frac{1}{12}$$

Reptiendo el cálculo visto para todos los elementos del espacio muestral conjunto, podemos hacer la asignación de probabilidades:

| $P$ | 1                | 2                | 3                | 4                | 5                | 6                |
|-----|------------------|------------------|------------------|------------------|------------------|------------------|
| $c$ | $P(c, 1) = 1/12$ | $P(c, 2) = 1/12$ | $P(c, 3) = 1/12$ | $P(c, 4) = 1/12$ | $P(c, 5) = 1/12$ | $P(c, 6) = 1/12$ |
| $x$ | $P(c, 1) = 1/12$ | $P(c, 2) = 1/12$ | $P(c, 3) = 1/12$ | $P(c, 4) = 1/12$ | $P(c, 5) = 1/12$ | $P(c, 6) = 1/12$ |

Puede advertirse fácilmente que la suma de todas las probabilidades es 1, como corresponde al suceso seguro.

### Pruebas repetidas de Bernoulli

Podemos pensar en experimentos compuestos por más de dos experimentos marginales. Un caso habitual es cuando se repite de forma **independiente**  $N$  veces un mismo experimento

$$\epsilon \times \dots \times \epsilon(\Omega \times \dots \times \Omega, \mathcal{F} \times \dots \times \mathcal{F}, P)$$

En este caso, podemos fácilmente asignar probabilidades:

$$P(A_1 \times \dots \times A_N) = P(A_1) \dots P(A_N)$$

Un caso habitual es de repetir **experimentos de Bernoulli**. Se trata de experimentos con dos resultados posibles, acertar o fallar, en lo que  $P(\text{acertar}) = p$  y  $P(\text{fallar}) = 1 - p$ .

Por ejemplo, en el lanzamiento de una moneda trucada la probabilidad de sacar cara (acertar, en este caso) es  $P(\text{cara}) = p$ .

Si repetimos  $N$  veces el experimento, ¿cuál será la probabilidad de acertar  $n$  veces?

Supongamos el siguiente suceso  $A$ , correspondiente a una secuencia de  $N = 10$  lanzamientos:

$$A = (c, c, x, x, x, c, c, x, c, c)$$

Es fácil ver que

$$P(A) = p^6(1 - p)^4$$

Cualquier secuencia que contenga 6 caras y 4 cruces tendrá idéntica probabilidad. En ocasiones, no nos interesa el orden en el que han salido los resultados, sino el número total de los mismos. Por ejemplo, nos interesa saber cuál es la probabilidad de que al lanzar  $N = 10$  veces

la moneda tengamos  $n = 6$  caras (aciertos). En tales casos, es necesario calcular el número de combinaciones que podemos obtener de  $N$  elementos agrupados de  $n$  en  $n$ :

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} \quad \binom{10}{6} = \frac{10!}{6!(10-6)!} = 210$$

Por lo que la probabilidad de 6 aciertos o caras será  $210p^6(1-p)^4$ . En general, la probabilidad  $P(n; N)$  de  $n$  aciertos en  $N$  intentos es:

$$P(n; N) = \binom{N}{n} p^n (1-p)^{N-n}$$

## Parametrización de espacios de probabilidad

Es frecuente que un espacio de probabilidad tenga sus probabilidades parametrizadas por un hecho o experimento externo al mismo.

Supongamos, por ejemplo, un test diagnóstico de cierta enfermedad. El test admite dos resultados posibles:

- $T^+$  = "el test resulta positivo" y
- $T^-$  = "el test resulta negativo",

Sin embargo, el resultado del test está íntimamente relacionado con la salud de la persona a la que se le aplica:

- $S_E$  = "SI tiene la enfermedad" y
- $N_E$  = "NO tiene la enfermedad".

El resultado del test se modela mediante un espacio de probabilidad. Ciertamente, el test puede fallar y resultar positivo cuando la persona está sana, o negativo cuando está enferma. Esto justifica que se modele probabilísticamente.

Sin embargo, es evidente que el modelado probabilístico debe tener en consideración si la persona a la que se le aplica el test está sana o enferma. De otro modo, el test sería inútil. Esperamos que el test de positivo cuando la persona esté enferma y negativo cuando está sana, y no al revés, sin perjuicio de que pueda equivocarse a veces.

Este razonamiento es propio de situaciones en las que existen **relaciones de causalidad que no son deterministas**: la enfermedad o ausencia de la misma es **causa del efecto observable** consistente en tener un resultado positivo o negativo en el test, pero el test puede fallar por lo que la relación no es determinista.

Esta preocupación se extiende a una infinidad de situaciones diferentes en múltiples campos muy distantes entre sí. Pensemos en un RADAR, que es un sistema diseñado para la detección y seguimiento de blancos, con distintas aplicaciones tanto civiles como militares. Un suceso que podemos medir es que el RADAR detecta un blanco, y queremos estar seguros de que realmente hay un blanco y de que no se trata de una *falsa alarma*. Otro suceso que podemos medir es que el RADAR no detecta nada, pero nuestro interés está en que realmente no haya blanco y no se haya producido una *pérdida*.

Otro ejemplo es el de un sistema de comunicación digital, que en cada instante transmite uno entre  $N$  símbolos posibles que tiene en su alfabeto. Sin embargo, al codificarlo en un señal electromagnética que, en su propagación, se ve afectada por distorsión y atenuación del canal, interferencias y ruido, pueden producirse errores en la detección. En este caso, los sucesos que podemos

medir son los símbolos que recupera el receptor, mientras que los sucesos relacionados en los que estamos interesados son los símbolos que realmente ha transmitido el emisor.

## Estrategia clásica versus bayesiana

Al modelar la relación del espacio de probabilidad observable con la causa subyacente, podemos considerar:

- **Estrategia clásica:** Que la causa es un parámetro determinista. Las verosimilitudes de las causas resultan de las probabilidades de las observaciones parametrizadas por las causas posibles:

$$- L(S_E) = P(T^+; S_E) = 1 - P(T^-; S_E)$$

$$- L(N_E) = P(T^-; N_E) = 1 - P(T^+; N_E)$$

- **Estrategia bayesiana:** Que la causa es, a su vez, modelable como un experimento aleatorio, en cuyo caso estaremos ante un experimento compuesto. Ahora **las versosimilitudes resultan de las probabilidades condicionadas de la observación por las causas**. Las probabilidades marginales corresponden a las probabilidades marginales de las causas, y **podemos calcular las probabilidades *a posteriori* de las causas, vistas las observaciones**, con el Teorema de Bayes.

$$- L(S_E) = P(T^+ / S_E) = 1 - P(T^- / S_E)$$

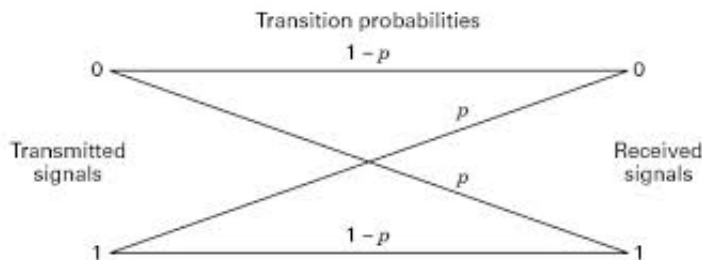
$$- L(N_E) = P(T^- / N_E) = 1 - P(T^+ / N_E)$$

Un caso de especial interés es aquel en el que solo hay **dos causas posibles** que identificamos como **hipótesis** a determinar a partir de las **observaciones**. Algunos ejemplos sencillos son

- El test de enfermedad, en el que tenemos dos posibles causas,  $S_E$  (hay enfermedad) y  $N_E$  (no hay enfermedad), y dos posibles efectos,  $T$  (test positivo) y  $T^-$  (test negativo).

Otros equivalentes son:

- La detección RADAR, donde las causas son "hay blanco" y "no hay blanco" y los efectos "blanco detectado" y "blanco no detectado".
- Un sistema de comunicación digital, en el que las causas son  $0_T$  (se transmite 0) y  $1_T$  (se transmite 1) y los efectos son  $0_D$  (se detecta 0 en el receptor) y  $1_D$  (se detecta 1 en el receptor)



Las hipótesis se refieren a las causas de las observaciones. Nuestra incertidumbre sobre las mismas puede modelarse como parámetros deterministas desconocidos o como sucesos aleatorios:

- **Hipótesis nula,  $H_0$** , se refiere al hecho o suceso de, por ejemplo, no existir enfermedad, blanco o nivel.
- **Hipótesis alternativa,  $H_1$** , corresponde a la situación complementaria.

Las observaciones mantienen una relación probabilística con las causas, bien en forma de verosimilitudes parametrizadas por las mismas, bien en forma de probabilidades condicionadas. Las hipótesis se seleccionan a partir de las observaciones, considerando su relación con las causas. Las observaciones son aleatorias, por lo que no determinan con certeza las causas:

- **Observación nula,  $O_0$** , aparentemente la observación respalda la hipótesis nula, por ejemplo, un test negativo lleva a presumir que no hay enfermedad, o una ausencia de señal RADAR que no hay blanco.
- **Observación alternativa,  $O_1$** , aparentemente a observación respalda la hipótesis alternativa.

En el **esquema clásico**, la elección de la hipótesis se hace dependiendo de qué verosimilitud sea mayor para una observación dada:

- Se observa  $O_0$
- $L(H_0) = P(O_0; H_0) > L(H_1) = P(O_0; H_1) \implies H_0$ . Lógicamente el test se diseña para que así sea.
- $L(H_1) = P(O_0; H_1) > L(H_0) = P(O_0; H_0) \implies H_1$
- Se observa  $O_1$
- $L(H_1) = P(O_1; H_1) > L(H_0) = P(O_1; H_0) \implies H_1$ . Lógicamente el test se diseña para que así sea.
- $L(H_0) = P(O_1; H_0) > L(H_1) = P(O_1; H_1) \implies H_0$ .

Las posibles transiciones y sus probabilidades asociadas, parametrizadas por las hipótesis, son:

- $P(O_0; H_0)$ , que corresponde a la probabilidad de aceptar correctamente la hipótesis nula, de rechazar la alternativa (no hay blanco, o enfermedad,...) de detectar correctamente el 0, o **nivel de confianza**. A veces se llama **especificidad**.
- $P(O_1; H_1)$ , que corresponde a la probabilidad de rechazar correctamente la hipótesis nula, de aceptar la alternativa (hay blanco, o enfermedad,...) de detectar correctamente el 1, o **potencia del test**. También se llama **probabilidad de detección**,  $P_D$ , o **sensibilidad**.
- $P(O_1; H_0) \equiv \alpha$ , que corresponde a la probabilidad de rechazar erróneamente la hipótesis nula, aceptar erróneamente la hipótesis alternativa, falso positivo, o **nivel de significación**,  $\alpha$ , **probabilidad de error de tipo I**,  $P_{\epsilon_1}$ , a veces llamada **probabilidad de falsa alarma**,  $P_{FA}$ .
- $P(O_0; H_1) \equiv \beta$ , que corresponde a la probabilidad de aceptar erróneamente la hipótesis nula, rechazar erróneamente la hipótesis alternativa, falso negativo o **probabilidad de error de tipo II**,  $P_{\epsilon_2}$ , a veces llamada **probabilidad de pérdida**,  $P_M$ .

Dado que sólo hay dos hipótesis posibles, es fácil advertir que:

- $P(O_1; H_0) \equiv \alpha = 1 - P(O_0; H_0) \implies P(H_0; H_0) = 1 - \alpha$
- $P(O_0; H_1) \equiv \beta = 1 - P(O_1; H_1) \implies P(H_1; H_1) = 1 - \beta$

Los **tests o contrastes de hipótesis** suelen diseñarse de modo que:

- La potencia  $1 - \beta$  sea máxima, esto es, que sea máxima la sensibilidad o probabilidad de detección.
- Que el nivel de significación  $\alpha$  se mantenga por debajo de un umbral o, equivalentemente, que la especificidad o nivel de confianza se mantengan suficientemente altos.

Veamos un primer ejemplo práctico. Supongamos que un sistema RADAR se diseña con las siguientes características:

- Probabilidad de detección:  $P_D = 0.99$
- Probabilidad de falsa alarma:  $P_{FA} \equiv \alpha = 0.15$

La hipótesis nula,  $H_0$ , es la ausencia de blanco y la hipótesis alternativa,  $H_1$ , es la presencia del mismo.

¿Cuáles son sus probabilidades de transición?

- $P(O_1; H_1) = P_D = 1 - \beta = 0.99$ , que corresponde a la **sensibilidad o potencia** del test.
- $P(O_1; H_0) \equiv \alpha = P_{FA} = 0.15$ , que corresponde al **nivel de significación** del test o probabilidad de error de tipo I.
- $P(O_0; H_1) \equiv \beta = 1 - P_D = 0.01$ , que es la probabilidad de pérdida y corresponde a la probabilidad de error de tipo II.
- $P(O_0; H_0) = 1 - \alpha = 1 - P_{FA} = 0.85$ , que corresponde al **nivel de confianza o especificidad** del test.

A modo de segundo ejemplo, consideremos un test clínico con las siguientes características:

- $H_0$ : no hay enfermedad
- $H_1$ : hay enfermedad
- Sensibilidad  $1 - \beta = 99\%$
- Especificidad:  $1 - \alpha = 85\%$

¿Cómo resultan ahora las probabilidades del test?

- $P(O_0; H_0) = 1 - \alpha = 0.85$  \*  $P(O_1; H_1) = 1 - \beta = 0.99$
- $P(O_1; H_0) \equiv \alpha = 0.15$  \*  $P(O_0; H_1) \equiv \beta = 0.01$

En el **esquema bayesiano**, se dispone de las probabilidades *a priori* de las hipótesis,  $P(H_0)$  y  $P(H_1)$  y de las probabilidades de transición condicionadas, o verosimilitudes:

- $L(H_0) = P(O_0/H_0) = 1 - P(O_1/H_0)$ ,
- $L(H_1) = P(O_1/H_1) = 1 - P(O_0/H_1)$

La **probabilidad total** de cada observación es:

$$P(O_0) = P(O_0/H_0)P(H_0) + P(O_0/H_1)P(H_1) \quad P(O_1) = P(O_1/H_0)P(H_0) + P(O_1/H_1)P(H_1)$$

Las probabilidades *a posteriori* resultan:

- Probabilidades de acierto:

$$P(H_0/O_0) = \frac{P(O_0/H_0)P(H_0)}{P(O_0)}P(H_1/O_1) = \frac{P(O_1/H_1)P(H_1)}{P(O_1)}$$

- Probabilidad total de acierto:

$$P(A) = P(H_0/O_0)P(O_0) + P(H_1/O_1)P(O_1)$$

- Probabilidades de error:

$$P(H_1/O_0) = \frac{P(O_0/H_1)P(H_1)}{P(O_0)}P(H_0/O_1) = \frac{P(O_1/H_0)P(H_0)}{P(O_1)}$$

- Probabilidad total de error:

$$P(E) = P(H_1/O_0)P(O_0) + P(H_0/O_1)P(O_1)$$

Podemos extender el número de hipótesis a un número  $K > 2$ . Esto resultará de utilidad en problemas de clasificación:

Las hipótesis  $\mathbf{H}_0 \dots \mathbf{H}_{K-1}$  corresponden a cada una de las clases (o causas) que subyacen como sucesos condicionantes, o que medimos.

Las probabilidades de transición son

- $P(O_0/H_0), P(O_1/H_0), \dots, P(O_{K-1}/H_0)$
- .....
- $P(O_0/H_{K-1}), P(O_1/H_{K-1}), \dots, P(O_{K-1}/H_{K-1})$

donde:

- $P(O_i/H_i), i = 0 \dots K - 1$  representan probabilidades de clasificación correcta.
- $P(O_i/H_j), i \neq j$  representan probabilidades de error.