A. ANEXO

Algunos aspectos utilizados en las referencias y los ejercicios a lo largo del documento hacen uso de principios estadísticos de uso general y que aquí se formalizan y se analizan en detalle para dotar de mayor rigor al documento.

Los sucesos pueden ser independientes o dependientes (condicionados, algunas denominaciones indican probabilidad relacionada).

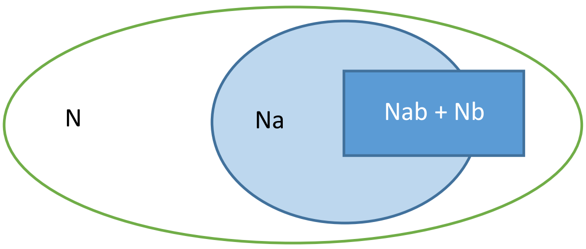
Está relacionado con el teorema de la probabilidad total y es de aplicación en los estudios de calidad y fallos identificando los fallos de tipo I (fallos identificados como fallo y que realmente lo son) y fallos de tipo II o falsos positivos (fallos consecuencia del modelo y del test utilizado y que realmente no existen). Fallos que el test identifica como tal y que realmente no lo son dado que ningún test tiene una fiabilidad absoluta (100% fiable que entonces sería un modelo o ley de ocurrencia). Se expone a continuación un resumen obtenido de Sixto Ríos (1973), al cual se dirige al lector para mayor detalle.

A.1 PROBABILIDAD CONDICIONADA, SUCESOS INDEPENDIENTES

Si se tiene un suceso A de probabilidad de ocurrencia P(A)>0, se entiende por probabilidad del suceso B condicionada por el suceso A, a la probabilidad de que ocurra el suceso B supuesto que haya ocurrido el suceso A, hecho que se denota P(B/A).

La probabilidad del suceso B condicionada por A ha de expresarse por una medida de en qué forma B está contenido en A (implicado por A), que ha de ser relativa a la probabilidad de A. Así, si en N tests se ha obtenido Na veces el suceso A y de entre estas, ha resultado Nab veces el B, (el total de sucesos B es Nab+Nb, Nb sucesos de B cuya unión con A es nula) se tiene:

$$fr(A)= Na /N; \quad f(B/A) = Nab /Na; \quad fr(A\cap B)= Nab /N$$



Como Nab /N = (Nab /N)·(Na/Na) esto es correcto dado que por definición es P(A) >0 o lo que es igual Na>0, luego se tiene:

$$fr(A\cap B) = Nab /N = (Nab /N)\cdot(Na/Na) = (Nab / Na)\cdot (Na /N) = fr(B/A)\cdot fr(A) \rightarrow$$

$$fr(B/A)=fr(A\cap B)/fr(A),$$

Pasando de las frecuencias a las probabilidades, en un espacio probabilístico tal que P(A)>0, se obtiene la expresión para la probabilidad condicionada de un suceso B por la ocurrencia de otro suceso A:

$$P(B/A) = P(A \cap B) / P(A) \qquad (A.1)$$

Si no existen sucesos comunes Nab=0, luego $P(A \cap B) = 0$, con lo que la expresión anterior P(B/A)=0; sin embargo esto no implica que ésta sea la única condición de independencia ya que pueden tener lugar sucesos (B) incluso en coexistencia con sucesos (A), pero que las probabilidades no dependan de esta circunstancia.

Sucesos dependientes e independientes. En general si $P(B/A) \neq P(B)$, en este caso B depende de A y si P(B/A) = P(B), B es independiente de A. En este último caso se tiene:

$$P(B/A) = P(B) = P(A \cap B) / P(A) \ \rightarrow \ P(B) \cdot P(A) = P(A \cap B) \qquad (A.2)$$

Y si B no depende de A entonces A no depende de B, en efecto; si B no depende de A se tiene:

$$P(A/B) = P(A \cap B) / P(B) = \text{aplicando (A.2)} = P(B) \cdot P(A) / P(B) = P(A)$$

Luego para la intersección de dos sucesos independientes vale la fórmula simétrica (A.2).

Teorema de la probabilidad total. Sean A_1, A_2, A·,... A_n, n sucesos mutuamente excluyentes y que forman un sistema exhaustivo, es decir, tales que:

$A_1 \cup A_2 \cup A_3 \cup \ldots \ A_n = E$, y sea B un suceso para el que se conocen las probabilidades condicionales P(B/Ai), y se conocen las probabilidades $P(A_i)$, se tiene:

$$P(B) = P(B \cap E) = P[B \cap (\cup A_i)] = \sum P(B \ \cap A_i) = \sum P(A_i) \cdot P(B|A_i),$$

$$P(B) = \sum P(A_i) \cdot P(B|A_i) \qquad (A.3)$$

Esta expresión es el teorema de la probabilidad total. El cual, utilizando las frecuencias de ocurrencia de una muestra representativa (suficientemente grande) y extendiéndolo al campo de las probabilidades[1], es el usado normalmente para el cálculo de la tasa de fallos, valor (l) de las funciones de probabilidad de Poisson.

Se obtiene de lo anterior, las relaciones siguientes:

$P(A_i \cap B) = P(A_i)\ P(B/A_i) = P(B)\ P(A_i/B)$, luego

$P(A_i/B) = P(A_i)\ P(B/A_i)\ /\ P(B)$, y sustituyendo ahora la expresión (A.3) se obtiene la fórmula de Bayes:

---

[1] El paso de frecuencia a probabilidades requiere del uso adecuado de los principios de la estadística y del concepto de probabilidad e intervalos de confianza. Si se utiliza de forma directa se está asumiendo el principio de trabajar con valores medios de ocurrencia lo que da valores aceptables para la programación del servicio de mantenimiento.

$$P(A_i/B) = P(A_i)\, P(B/A_i) \,/\, P(B) \rightarrow P(A_i/B) = P(A_i)\, P(B/A_i) \,/\, \sum P(A_i) \cdot P(B|A_i) \quad (A.4)$$

Ejercicio A1. Aplicación directa del teorema. Modificado de Ríos (1973).

Se dispone de tres urnas (A_1, A_2, A_3) con las configuraciones siguientes: A_1, 3 bolas blancas y 1 negra; A_2, 2 blancas y 2 negra; A_3, 1 blanca y 3 negras. De ellas se toma una urna al azar de la cual se saca una bola al azar. Indicar la urna que da una probabilidad mayor al suceso obtener una bola blanca (suceso B).

Solución:

$P(A_1) = P(A_2) = P(A_3) = 1/3$,

$P(B/A_1) = 3/4$; $P(B/A_2) = 2/4$; $P(B/A_3) = 1/4$,

Luego:

$P(A_1/B) = (1/3 \cdot 3/4) /\{(1/3)(3/4 + 2/4 +1/4)\} = 3/6$,

Análogamente:

$P(A_2/B) = 2/6$; $P(A_3/B) = 1/6$,

Luego la urna A_1 tiene la mayor probabilidad.

Ejercicio A2. Aplicación al control de calidad.

Ejercicio B1. Problema de control de calidad

A) En una fábrica se producen tornillos que se empaquetan en cajas de 1200 Ud. La experiencia, estudios de control de calidad, nos dan la probabilidad de que una determinada caja tenga un número de tornillos defectuosos según la tabla siguiente.

% tornillos defecuosas	% de cajas	Probab. defecto	Probabili-dad caja	Defectos por caja max.
0	78,0	0,000	0,780	0
1	17,0	0,010	0,170	12
2	3,4	0,020	0,034	24
3	0,9	0,030	0,009	36
4	0,5	0,040	0,005	48
5	0,2	0,050	0,002	60
6	0,0	0,060	0,000	72

B) La inspección se realiza analizando una muestra de 50 tornillos de cada caja. Una determinada caja da 6 defectuosos en los 50 inspeccionados. Analizar cuál es la probabilidad de que sea considerada correcta. Se considera una caja correcta si tiene un porcentaje de defectos menor o igual del 2%.

Suceso Ai: Escoger una caja

Suceso Bi: Que tenga un determinado nivel de defecto

A) Probabilidad total de defecto, P(B): 0,00295 — P(B)=∑P(Ai)·P(Bi|Ai)

Probabilidad de caja con defectos ≤ 2%, y >2% — Pr(Ai|Bi)=P(Ai)·P(Bi|Ai)/P(B)

P(B0):	0,0000			
P(B1):	0,5763			
P(B2):	0,2305	Prob(B≤2)=	0,8068	=suma{P(B0)+P(B1)+P(B2)}
P(B3):	0,0915			
P(B4):	0,0678			
P(B5):	0,0339			
P(B6):	0,0000	Prob(B>2)=	0,1932	P(B≤ + B>)= 1,0000

Ejercicio A3. Falso positivo.

La probabilidad de fallo (tornillo roto) en la fabricación de tornillos es de 1/10000. La probabilidad de identificar un positivo (roto) con el test es del 99% (P(positivo|roto)= 0,99) y la probabilidad de que se dé un falso positivo, error en la identificación, es del (P(positivo|íntegro))[2]=0,05.

Determinar:

a) La probabilidad de que cuando el tornillo esté roto se acierte en la identificación.

b) Probabilidad de que cuando un tornillo esté roto se identifique después de un doble test.

c) Ha salido positivo el test, ¿Cuál es la probabilidad de que esté roto?

d) Probabilidad de que esté roto tras un doble test positivo.

Solución:

a) Aplicación del teorema de Bayes o de la probabilidad compuesta:

$$P(B|A)=P(A|B)\cdot(P(B)/P(A)$$

La probabilidad de que el tornillo esté roto es de:

$$P(roto) = P(e) = 0{,}0001$$

La probabilidad de que cuando tornillo esté roto se acierte en la identificación:

$$P((positivo|roto) = P(p|r) = 0{,}99$$

b) Si ahora se hace un segundo test para verificar el resultado, la probabilidad de error en ambos casos, son sucesos independientes el $test_1$ y el $test_2$, es:

$P(error, test_1+test_2)=(1-0{,}99)\cdot(1-0{,}99)=0{,}0001$; luego la probabilidad de acertar es:

$$P(roto, test_1+test_2)=1-0{,}0001 = 0{,}9999 \rightarrow 99{,}99\%$$

El test doble es muy fiable para identificar un roto cuando realmente lo es.

c) La probabilidad de falso positivo (error tipo II) es:

$$P(positivo|integro) = P(p|i) = 0{,}05$$

[2] El suceso positivo|roto, no es complementario al suceso positivo|íntegro. Los sucesos complementarios son (positivo|roto, con negativo|roto) y (positivo|íntegro con negativo|íntegro); por esta razón no suma 1 los valores P(positivo|roto)=0,99 más P(positivo|íntegro)=0,05, cuya suma es 1,04. En la realidad la fiabilidad de un test para detectar una enfermedad no tiene por qué ser complementario a la de error tipo II (dar un falso positivo), y de hecho no lo es como norma salvo coincidencia. Este hecho es causa de errores importantes que se deben evitar. Hay test que dan un claro positivo hasta un valor, un claro negativo hasta otro valor y una zona de duda o incertidumbre de interpretación.

Ha salido positivo el test, ¿Cuál es la probabilidad de que realmente está roto?

P(roto|positivo)=P(r|p)[3]= P(positivo|roto)·P(roto) / P(positivo)

P(r|p) = P(p|r) · P(r) /P(p) = P(p|r)·P(r) / {P(r)·P(p|r) + P(i)·P(p|i)

Sustituyendo valores se tiene:

P(r|p) = 0,99·0,0001 / (0,0001·0,99 + 0,9999 · 0,05) = 0,002 →0,2 %

d) Si ahora repetimos la prueba, dado que ha salido positivo y queremos verificar el resultado, se tienen dos sucesos excluyentes el primer test y el segundo test, los resultados serán

Prueba test 1: $P(test_{1positivo}) = 0,002$

Prueba test 2: $P(test_{2positivo}) = 0,002$

El error en ambos test, por ser independientes, será el producto de probabilidades

$P(test_1 + test_2) = 0,002 \cdot 0,002 = 0,000004 = 0,0004\%$

Esta probabilidad conjunta es tan baja que debemos pensar que no se da realmente, luego si en ambos test sale positivo tenemos una probabilidad, de que sea cierta muy alta, del 99,96% calculado sobre el error del test y del 99,99% sobre la certeza del test; luego se debe concluir que en efecto el resultado es correcto.

Lo anterior, aplicado como ejercicio al supuesto de la rotura de un tornillo es un mero ejercicio técnico, pero aplicado realmente a un test de detección del cáncer es de extrema importancia, como así se considera realmente, que se duplica el test como principio de trabajo. En el ejemplo de la detección de enfermedades se identifica con mayor fuerza el problema del falso positivo, o sea que te identifiquen un cáncer cuando en realidad no lo tienes o la opción contraria, que no se detecte cuando en realidad ya se está desarrollando la enfermedad.

BIBLIOGRAFÍA

Ríos, Sixto. (1973). Métodos estadísticos. Ediciones del Castillo, España.