

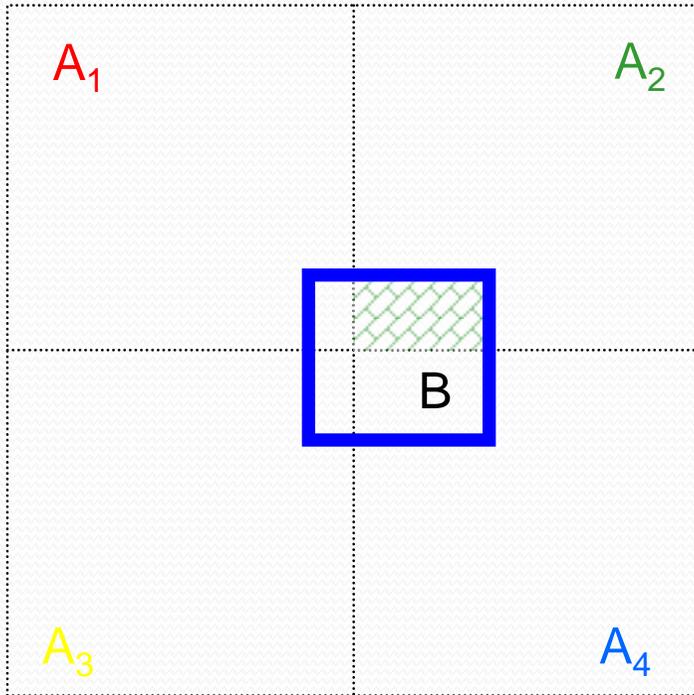
Estadística General

Tema 3: Probabilidad
Teorema de Bayes y Prueba diagnóstica



Prof. José G. Páez

Teorema de Bayes



Si conocemos la probabilidad de B en cada uno de los componentes de un sistema exhaustivo y excluyente de sucesos, entonces...

...si ocurre B, podemos calcular la probabilidad (*a posteriori*) de ocurrencia de cada A_i .

$$P(A_i | B) = \frac{P(B \cap A_i)}{P(B)}$$

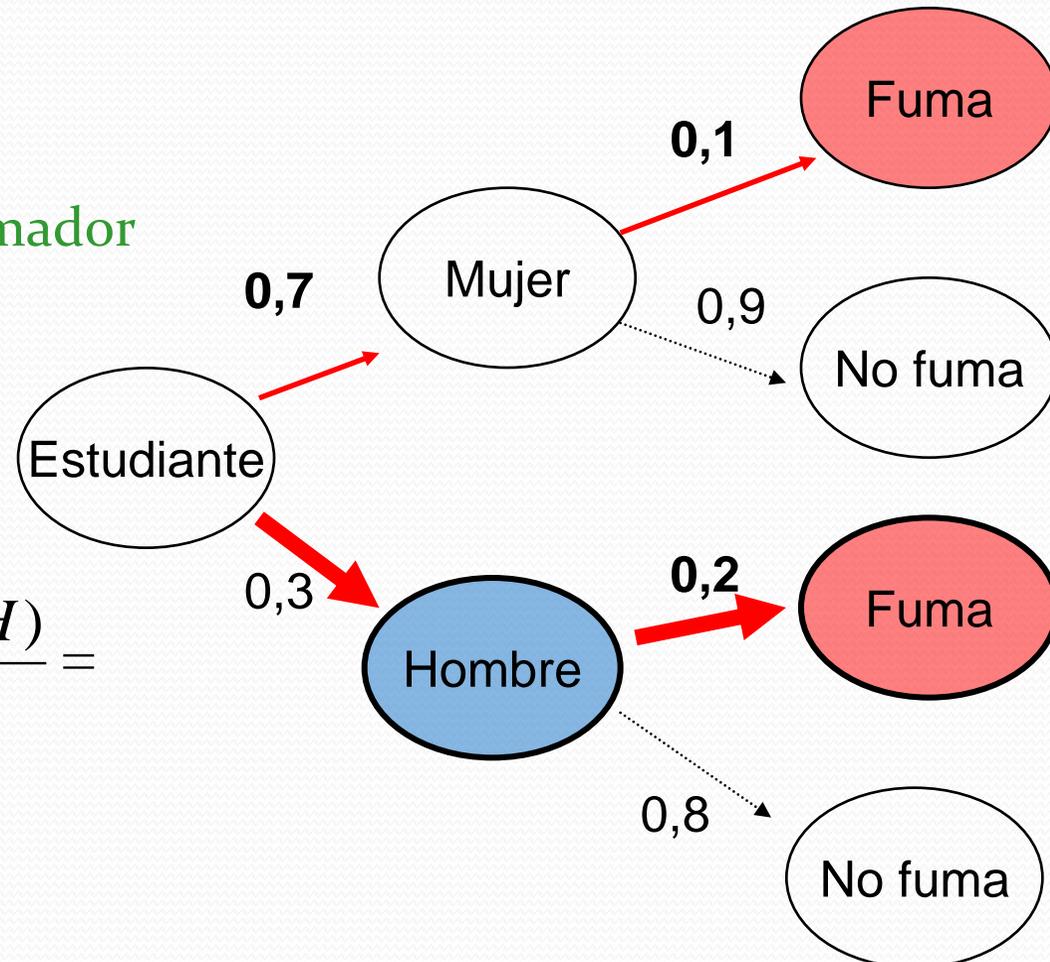
donde $P(B)$ se puede calcular usando el teorema de la probabilidad total:

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3) + P(B \cap A_4)$$

$$= P(B|A_1) P(A_1) + P(B|A_2) P(A_2) + \dots$$

Ejemplo (II): En este aula el 70% de los alumnos son mujeres. De ellas el 10% son fumadoras. De los varones, son fumadores el 20%.

- ¿Qué porcentaje de fumadores hay?
 - $P(F) = 0,7 \times 0,1 + 0,3 \times 0,2 = 0,13$
 - (Resuelto antes)
- Se elije a un individuo al azar y es... fumador
¿Probabilidad de que sea un hombre?

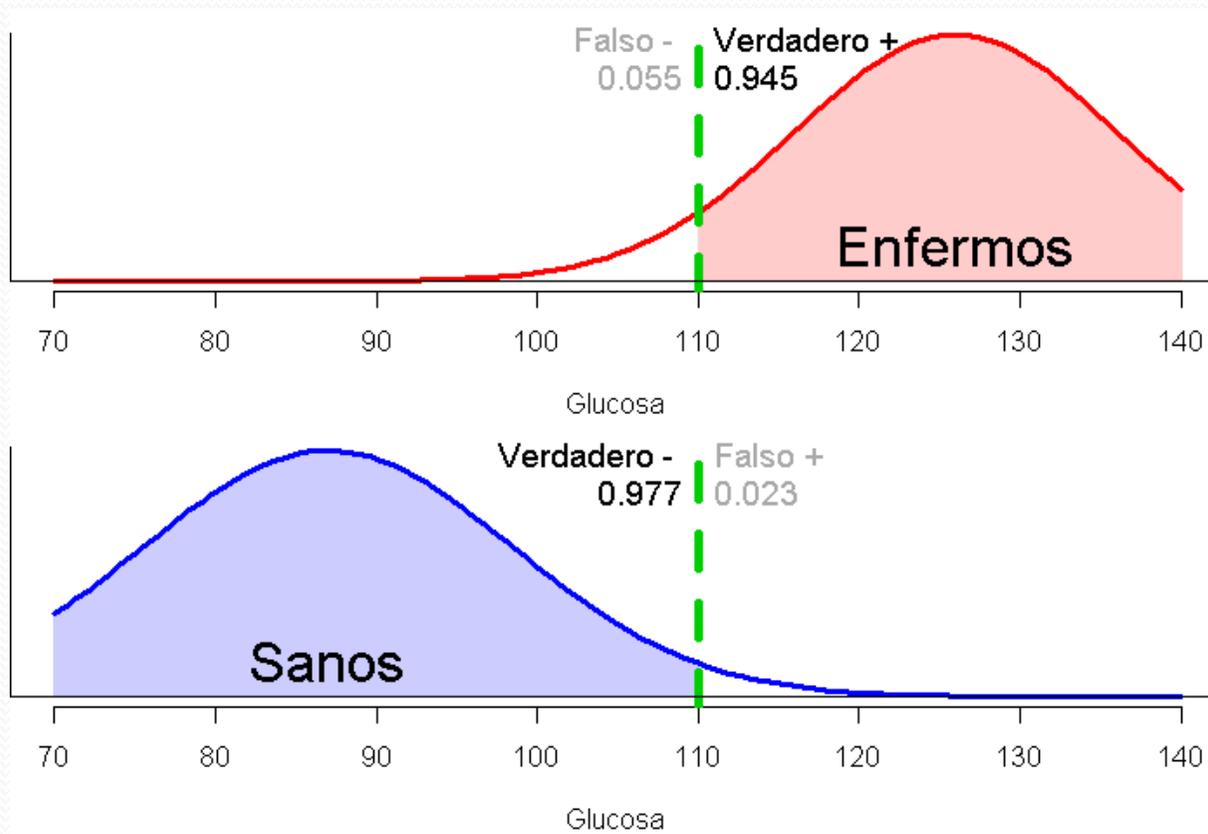


$$P(H | F) = \frac{P(H \cap F)}{P(F)} = \frac{P(H) \cdot P(F | H)}{P(F)} =$$
$$= \frac{0,3 \times 0,2}{0,13} = 0,46$$

Ejemplo de prueba diagnósticas: Diabetes

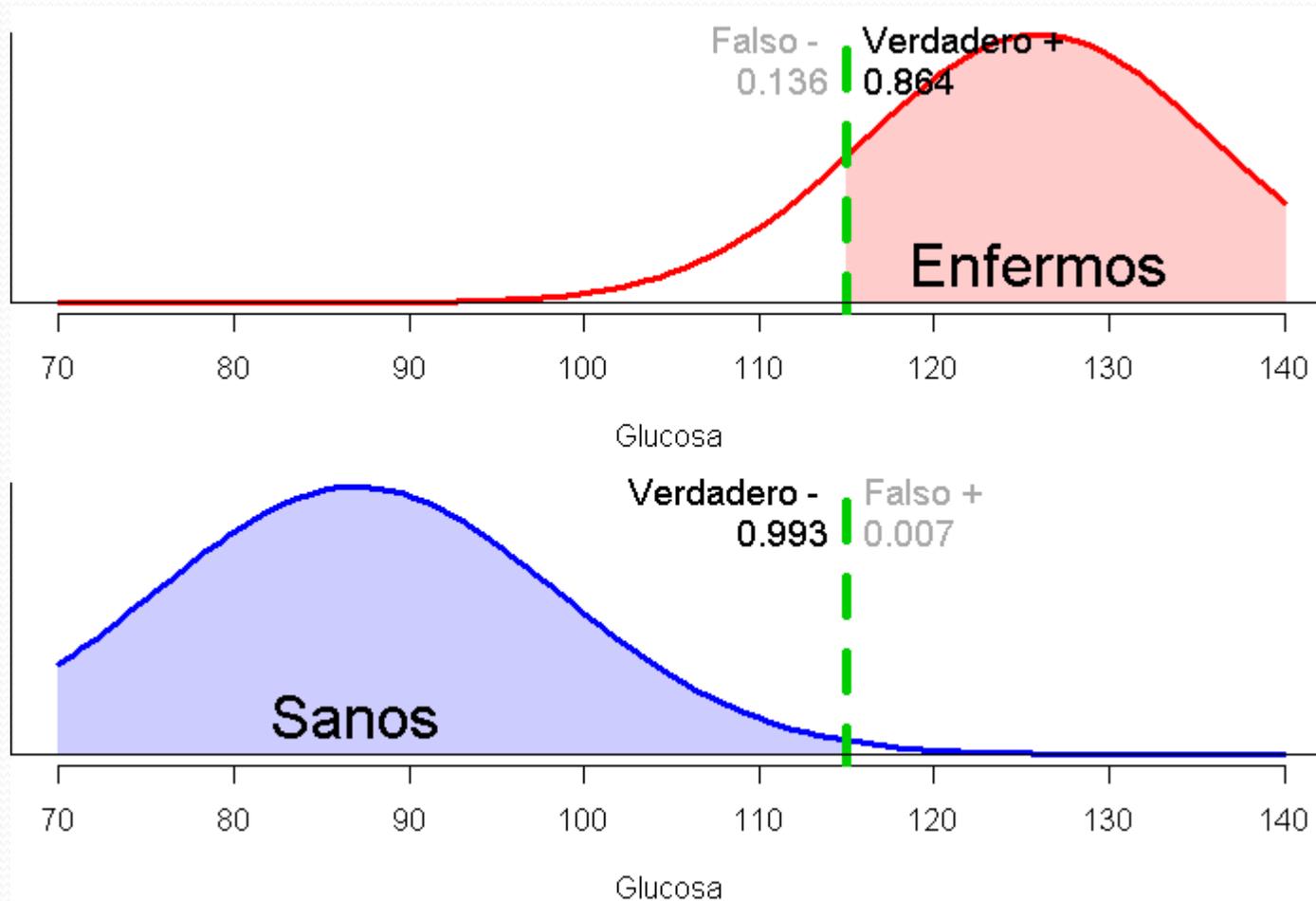
- Los carbohidratos ingeridos terminan como glucosa en la sangre. El exceso se transforma en glucógeno y se almacena en hígado y músculos. Este se transforma entre comidas de nuevo en glucosa según necesidades.
- La principal hormona que regula su concentración es la insulina. La diabetes provoca su deficiencia o bien la insensibilidad del organismo a su presencia. Es una enfermedad muy común que afecta al 2% de la población (**prevalencia**)
- Una prueba común para diagnosticar la diabetes, consiste en medir el nivel de glucosa. En individuos sanos suele variar entre 64 y 110mg/dL.
 - El cambio de color de un indicador al contacto con la orina suele usarse como indicador (**resultado del test positivo**)
- Valores por encima de 110 mg/dL se asocian con un posible estado pre-diabético.
 - Pero no es seguro. Otras causas podrían ser: hipertiroidismo, cancer de páncreas, pancreatitis, atracón reciente de comida...
- Supongamos que los enfermos de diabetes, tienen un valor medio de 126mg/dL.

Funcionamiento de la prueba diagnóstica de glucemia



- **Valor límite: 110mg/dL**
 - Superior: test positivo.
 - Inferior: test negativo.
- **Probabilidad de acierto:**
 - Para enfermos
 - Verdadero positivo (sensibilidad)
 - Para sanos
 - Verdadero negativo (especificidad)
- **Probabilidad de error**
 - Para enfermos
 - Falso -
 - Para sanos
 - Falso +

¿Cómo definir el punto de corte de la prueba diagnóstica?

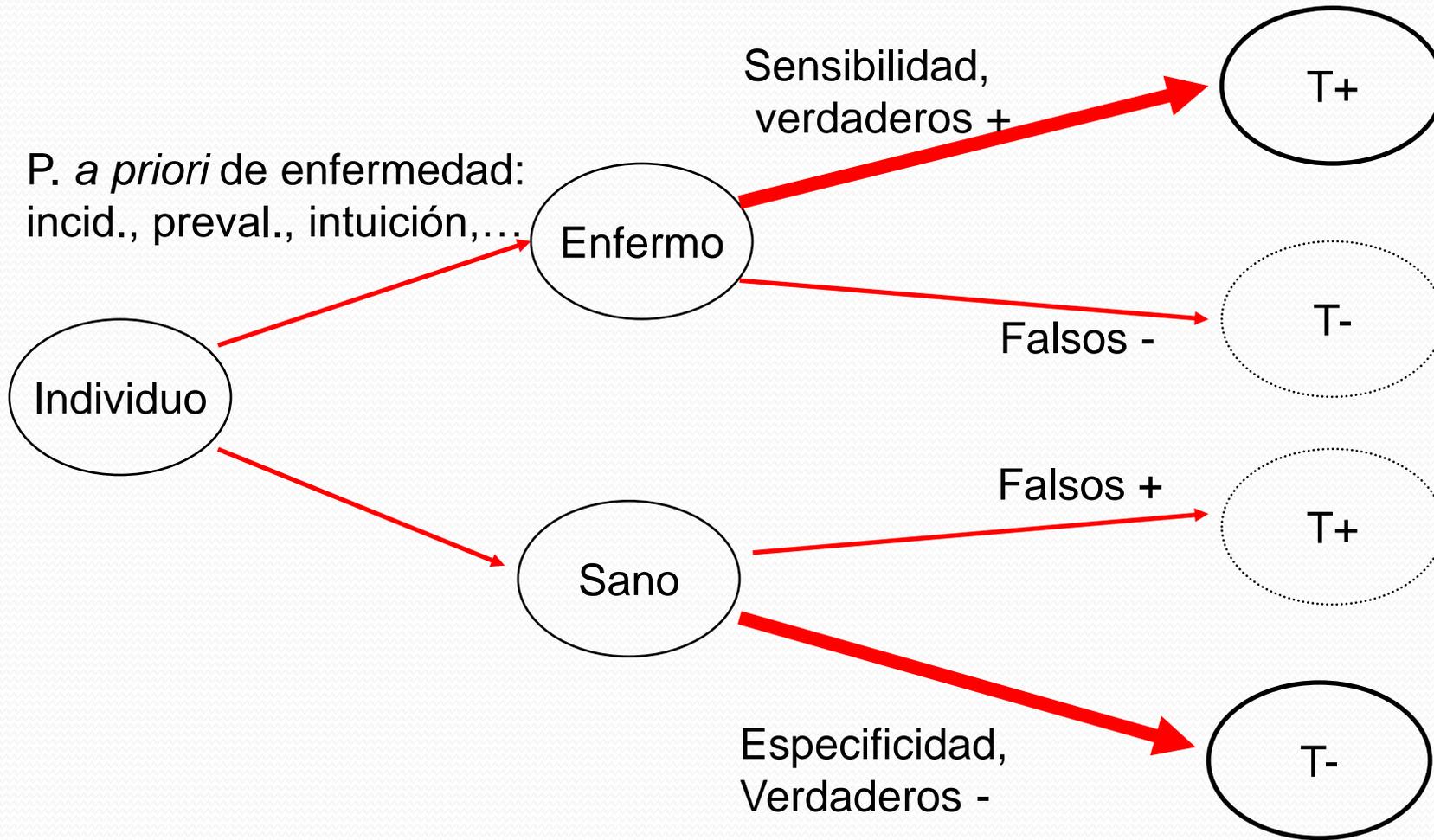


No es simple. **No es posible aumentar sensibilidad y especificidad al mismo tiempo.** Hay que elegir una solución de compromiso: Aceptable sensibilidad y especificidad.

Una **prueba diagnóstica** ayuda a mejorar una estimación de la probabilidad de que un individuo presente una enfermedad.

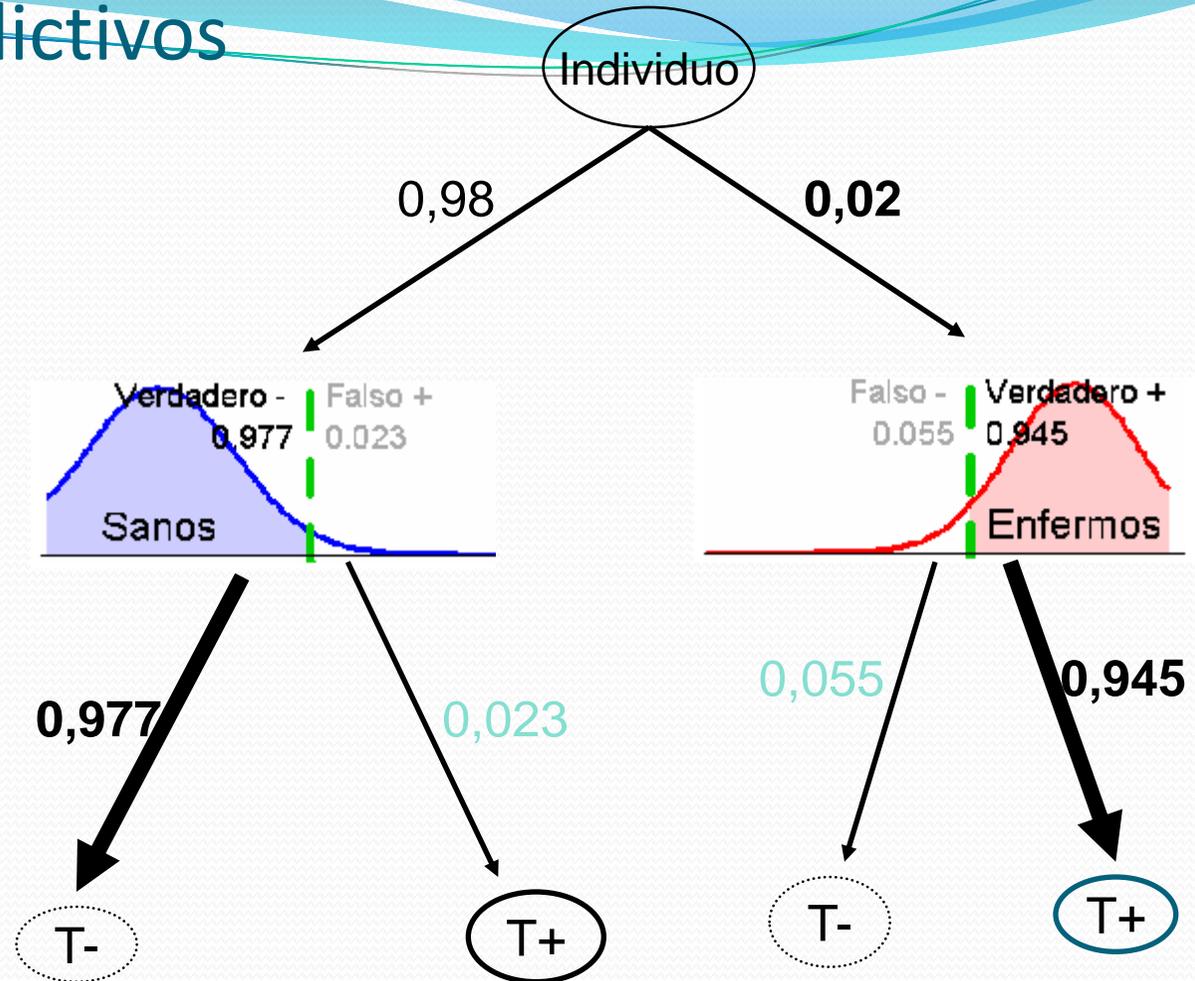
- En principio tenemos una **idea subjetiva** de $P(\text{Enfermo})$. Nos ayudamos de...
 - **Incidencia**: Porcentaje de nuevos casos de la enfermedad en la población.
 - **Prevalencia**: Porcentaje de la población que presenta una enfermedad.
- Para confirmar la sospecha, usamos una prueba diagnóstica. Ha sido evaluada con anterioridad sobre dos grupos de individuos: sanos y enfermos. Así **de modo frecuentista** se ha estimado:
 - $P(+ | \text{Enfermo}) = \text{Sensibilidad}$ (verdaderos +) = Tasa de acierto sobre enfermos.
 - $P(- | \text{Sano}) = \text{Especificidad}$ (verdaderos -) = Tasa de acierto sobre sanos.
- A partir de lo anterior y usando el **teorema de Bayes**, podemos calcular las probabilidades *a posteriori* (en función de los resultados del test): **Índices predictivos**
 - $P(\text{Enfermo} | +) = \text{Índice predictivo positivo}$
 - $P(\text{Sano} | -) = \text{Índice predictivo negativo}$

Pruebas diagnósticas: aplicación Teorema de Bayes.



Ejemplo: Índices predictivos

- La diabetes afecta al 2% de los individuos.
- La presencia de glucosuria se usa como indicador de diabetes.
- Su sensibilidad es de 0,945.
- La especificidad es de 0,977.
- Calcular los índices predictivos.



$$P(\text{Sano} | T-) = \frac{P(\text{Sano} \cap T-)}{P(\text{Sano} \cap T-) + P(\text{Enf} \cap T-)}$$

$$= \frac{0,98 \cdot 0,977}{0,98 \cdot 0,977 + 0,02 \cdot 0,055} = 0,999$$

$$P(\text{Enf} | T+) = \frac{P(\text{Enf} \cap T+)}{P(\text{Enf} \cap T+) + P(\text{Sano} \cap T+)}$$

$$= \frac{0,02 \cdot 0,945}{0,02 \cdot 0,945 + 0,98 \cdot 0,023} = 0,456$$

Observaciones

- En el ejemplo anterior, al llegar un individuo a la consulta tenemos una idea *a priori* sobre la probabilidad de que tenga una enfermedad.
- A continuación se le pasa una **prueba diagnóstica** que nos aportará nueva información: Presenta glucosuria o no.
- En función del resultado tenemos una nueva idea (*a posteriori*) sobre la probabilidad de que esté enfermo.
 - Nuestra opinión a priori ha sido modificada por el resultado de un experimento.

- ¿Qué probabilidad tengo de estar enfermo?

- En principio un 2%. Le haremos unas pruebas.



- Presenta glucosuria. La probabilidad ahora es del 45,6%.

