that helps you to predict any of (57)

the other I_j. Thus the I_i are

given θ, the I_i are $\boxed{\text{indep.}}$

unconditionally dependent but

conditionally independent given θ.

(4 Aug 17)

~~Bayes's Theorem for events~~ (a finite partition)

Suppose that the events $B_1, \dots, B_k$ partition the sample space in such a way that $P(B_j) > 0$ for all $j = 1, \dots, k$. If A is an event with $P(A) > 0$, then for each $i = 1, \dots, k$

$$P(B_i \mid A) = \frac{P(B_i)\, P(A \mid B_i)}{P(A)}$$

and, by the LTP, this is

$$P(B_i \mid A) = \frac{P(B_i) \cdot P(A \mid B_i)}{\sum_{j=1}^{k} P(B_j) \cdot P(A \mid B_j)}.$$

How this theorem is used in Bayesian statistics: the B_i represent unknown states of the world: they're all possible — $P(B_i) > 0$ — and only one of them is true, but you don't know which one. A represents <u>data</u>: information that will help you identify the most probable B_i.

Before (a priori) the dataset A arrives, you have background information about the plausibility of the B_i that you can represent with prior probabilities $P(B_i)$. After (a posteriori) the dataset A arrives, you can use Bayes's Theorem to update your prior probabilities to posterior probabilities $P(B_i \mid A)$.

The probabilities $P(A \mid B_i)$ represent how likely the dataset A would be if B_i were the actual unknown state; this is often called likelihood information.

(the denominator)
$P(A)$ does not depend on the B_i, and can therefore be regarded as a <u>normalizing constant</u>, put into Bayes's Theorem to make all the $P(B_i \mid A)$ add up to 1. Thus

$$P(B_i \mid A) = \frac{P(B_i)\, P(A \mid B_i)}{P(A)}$$

is interpreted as

$$\begin{pmatrix} \text{posterior} \\ \text{information} \end{pmatrix} = \frac{\begin{pmatrix} \text{prior} \\ \text{information} \end{pmatrix} \cdot \begin{pmatrix} \text{likelihood} \\ \text{information} \end{pmatrix}}{\begin{pmatrix} \text{normalizing} \\ \text{constant} \end{pmatrix}}.$$

Random variables and their distributions

Example: Tay-Sachs disease

T = T-S baby
N = not

$\mathcal{S}$	Y
NNNNN	0
TNNNN	1
NTNNN	1
NNTNN	1
NNNTN	1
NNNNT	1
TTNNN	2
TNTNN	2
TNNTN	2
TNNNT	2
NTTNN	2
NTNTN	2
NTNNT	2
NNTTN	2
NNTNT	2
NNNTT	2
⋮	⋮
TTTTT	5

← # of T-S babies = Y

Definition

Given a sample space $\mathcal{S}$ for an experiment $\mathcal{E}$, a (real-valued) random variable (rv) is a function from the non-weird collection $\mathcal{C}$ of subsets of $\mathcal{S}$ to the real number line $\mathbb{R}$.

In the T-S case study, the elements s of $\mathcal{S}$ look like NNNTN and the rv Y counts how many Ts they contain.

For instance, $Y(TNNTN) = 2$ and $Y(NNNTT) =$ (also) 2 (i.e., Y ignores the order of the children). We can use the following notation to simplify things.

Notation: $\underbrace{P(Y = y)}_{\text{T/F proposition}} \doteq P(\overbrace{\{s : Y(s) = y\}}^{\text{set}})$

For example, $P(Y = 1) = P(\{s \in S : Y(s) = 1\})$

$= P(\{TNNNN, NTNNN, NNTNN, NNNTN, NNNNT\})$.

In general the values a random variable takes on could be just about anything, but in this course all of our rvs will be real-valued.

In the T-S case study the rv Y can only take on the values $0, 1, \dots, 5$.

(7 Aug 17)