Statistics

Trong Truong

November 29, 2018

1 Conditional probability

If the conditional probability of A given B is equal to the (unconditional) probability of B, we say that A and B are statistically independent. This means occurrence of B does not change the probability of A occurring. This means that

$$
\begin{aligned}
P(A) = P(A|B) &= \frac{P(A \cap B)}{P(B)} \\
P(A)P(B) &= P(A \cap B) \quad \text{(Law of multiplication).}
\end{aligned}
$$

2 Mendel's Laws of Heredity

It is amazing that you can deduce a law about the *discrete* nature of heredity (genes) from the statistics of plant breeding, specifically from the proportion

$$0.749 \approx 0.75 = \frac{3}{4}$$

of offspring of hybrid peas showing dominant traits. Whenever such neat ratios are found in nature, they almost surely indicate a discreteness in the underlying phenomenon.

3 General Law of Addition / Inclusion Exclusion Principle

Theorem 1 (Inclusion Exclusion Principle)**.** *Let $E_1, \ldots, E_n$ be any events. Then*

$$P(E_1 + \cdots + E_n) = \sum_i P(E_i) - \sum_{i<j} P(E_i E_j) + \sum_{i<j<k} P(E_i E_j E_k) - \cdots (-1)^{n-1} P(E_1 \cdots E_n).$$

Proposition 2. *Let $E_1, \dots, E_n$ be any events. Then*

$$\sum_i P(E_i) - \sum_{i<j} P(E_i E_j) \leq P(E_1 + \cdots + E_n) \leq \sum_i P(E_i).$$

Note 3. Note that $E_1 + \cdots + E_n$ is the event where at least one of the E_i's occurs.

Proof. This is obvious if we think of the events E_i as regions in a plane and P as the area function. For example, in this case $P(E_1 + \cdots + E_n)$ represents the area A covered by all the E_i's; $\sum_i P(E_i) - \sum_{i<j} P(E_i E_j)$ represents the area covered by all the E_i's except the intersections of any pair E_i, E_j, which is clearly a smaller area than A. □

Definition 4 (Pairwise and mutual independence)**.** The events $E_1, \dots, E_n$ are called pairwise independent if

$$P(E_i E_j) = P(E_i) P(E_j)$$

for any pair $i \neq j$. They are called mutually independent if they also satisfy

$$\begin{aligned} P(E_i E_j E_k) &= P(E_i) P(E_j) P(E_k) \\ &\cdots \\ P(E_1 \cdots E_n) &= P(E_1) \cdots P(E_n) \end{aligned}$$

for any set of distinct $i, j, \dots$.

Example 5. Pairwise independence does not imply mutual independence. E.g. Consider the experiment of throwing two dice, one red and one black, and let E_1, E_2, E_3 be the events of throwing odd on red, odd on black, and an odd sum on both dice, respectively. Then it's easy to verify that E_i are pairwise independent, but they are not mutually independent because

$$P(E_1 E_2 E_3) = 0 \neq \frac{1}{8} = P(E_1) P(E_2) P(E_3).$$

4 Random Variables and Probability Distribution