Class Notes 1-28-2019

- **Bayes' Formula**:

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)} \quad \text{for } P(A), P(B), P(B^c) > 0$$

- **Example: Blood Test for Disease**
A test is 95% effective on persons with the disease and has a 1% false alarm rate. Suppose that the prevalence of the disease in the population is .5%. What is the chance that the person actually has the disease (event D), given that the test is positive (event E)?

$$\begin{aligned} P(D|E) &= \frac{P(DE)}{P(E)} = \frac{P(E|D)P(D)}{P(E|D)P(D) + P(E|D^c)P(D^c)} \qquad \text{Bayes' formula} \\ &= \frac{.95 \cdot .005}{.95 \cdot .005 + .01 \cdot .995} = \frac{95}{294} \approx .323 \end{aligned}$$

With $P(E|D) = .99$

$$P(D|E) = \frac{.99 \cdot .005}{.99 \cdot .005 + .01 \cdot .995} \approx .3322$$

With $P(E|D) = 1$

$$P(D|E) = \frac{1 \cdot .005}{1 \cdot .005 + .01 \cdot .995} \approx .33445$$

The following long run type argument makes the surprising answer more transparent. Out of 1000 people, roughly 995 will have no disease, and about 10 of them will give a false positive E.
5 will have the disease and about all will give a true positive. $5/(10 + 5) \approx .333$.
Such illustrations can counter the possible psychological damage arising from routine tests. Of course, a doctor's interpretation of the presented patient symptoms may justify a much higher prevalence of the disease among such patients than the above .5%.

- **Generalized Bayes' Formula**: Let $B_1, \ldots, B_n$ be a partition of Ω, i.e., $B_i \cap B_j = \emptyset$ for all i, j and $B_1 \cup \ldots \cup B_n = \Omega$. Then for $P(B_i) > 0$ for all i and $P(A) > 0$ we have for any k

$$P(B_k|A) = \frac{P(AB_k)}{P(A)} = \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^n P(A|B_i)P(B_i)}$$

- **Example: Search**
Suppose you have mislaid an important item and you have 3 general locations where it might be. Assume for simplicity that all 3 locations have the same probability 1/3 of containing the item. Let A_i be the event that the item is in location i, $i = 1, 2, 3$. These locations present different difficulties in finding the item. Denote by E the event that search of location $i = 1$ is unsuccessful. Denote by

$$\beta_i = P(E|A_i) = \text{ the probability of missing the item in location 1 when it is in location } i$$

Find the probabilities that the item is in location i , $i = 1, 2, 3$, given that a search of location 1 was unsuccessful.

$$P(A_1|E) = \frac{P(E|A_1)P(A_1)}{P(E|A_1)P(A_1) + P(E|A_2)P(A_2) + P(E|A_2)P(A_3)} = \frac{\beta_1}{\beta_1 + 1 + 1} = \frac{\beta_1}{\beta_1 + 2}$$

while for $j = 2, 3$ we have

$$P(A_j|E) = \frac{P(E|A_j)P(A_j)}{P(E|A_1)P(A_1) + P(E|A_2)P(A_2) + P(E|A_2)P(A_3)} = \frac{1}{\beta_1 + 1 + 1} = \frac{1}{\beta_1 + 2}$$

- **Independence:** Two events are **independent** if and only if

$$P(AB) = P(A)P(B)$$

Motivate through $P(B|A) = P(B)$, but it requires $P(A) > 0$ and it seem asymmetric in A and B. If both $P(A)$ and $P(B)$ are > 0 then $P(B|A) = P(B) \Leftrightarrow P(A|B) = P(A)$. Our definition of independence actually implies independence when $P(A) = 0$ or $P(B) = 0$. Why?

- **Example: 2 Dice**
Roll two fair dice. The number on the first die is independent of the number on the second die. Verify this using events describing outcomes of two dice separately, like even on first die, E, and divisible by 3 on second die, D_3. $P(ED_3) = 6/36 = P(E)P(D_3) = 3/6 \cdot 1/3$
Now let F_4 be the event that the first die is 4 and S_6 the event that the sum is 6 and S_7 the event that the sum is 7. Is F_4 independent of S_6 (S_7)?
$P(F_4) = \frac{1}{6}$, $P(S_6) = \frac{5}{36}$, $P(S_7) = \frac{1}{6}$

$$P(F_4S_7) = \frac{1}{36} = P(F_4)P(S_7) \ , \ P(F_4S_6) = \frac{1}{36} \neq P(F_4)P(S_6)$$

- **Example: Cards**
We draw a card at random from a standard deck of 52 cards. The event A that the card is an ace is independent of the event C that the card is a club. However, this breaks down as soon as the king of diamonds is missing from the deck, but not when all kings are missing.