Statistics 351 (Fall 2015) September 30, 2015
Prof. Michael Kozdron

Lecture #10: Conditioning (Chapter 2)

Reference. §2.1 pages 31–33

Suppose that A and B are events with $P\{A\} > 0$ and $P\{B\} > 0$. Since we can write A as the disjoint union $A = (A \cap B) \cup (A \cap B^c)$, we conclude that

$$P\{A\} = P\{A \cap B\} + P\{A \cap B^c\}. \tag{$*$}$$

In general, if $B_1, B_2, \ldots, B_m$ partition the sample space Ω, that is if $B_1, B_2, \ldots, B_m$ are disjoint with $B_1 \cup \cdots \cup B_m = \Omega$ and $P\{B_k\} > 0$ for all $k = 1, \ldots, m$, then

$$P\{A\} = \sum_{i=1}^{m} P\{A \cap B_i\}.$$

This formula is sometimes called the *law of total probability.* The *conditional probability* of A given B is defined to be

$$P\{A|B\} = \frac{P\{A \cap B\}}{P\{B\}} \tag{$**$}$$

so that, similarly, the conditional probability of B given A is

$$P\{B|A\} = \frac{P\{B \cap A\}}{P\{A\}}.$$

Since $P\{A \cap B\} = P\{B \cap A\}$ we conclude

$$P\{B|A\} = \frac{P\{A|B\}P\{B\}}{P\{A\}}. \tag{$***$}$$

Now, substituting $(**)$ into $(*)$ we find

$$\begin{aligned} P\{A\} &= P\{A \cap B\} + P\{A \cap B^c\} \\ &= P\{A|B\}P\{B\} + P\{A|B^c\}P\{B^c\} \end{aligned}$$

so that $(***)$ implies

$$P\{B|A\} = \frac{P\{A|B\}P\{B\}}{P\{A|B\}P\{B\} + P\{A|B^c\}P\{B^c\}}.$$

This result is sometimes known as *Bayes' rule.* In fact, it is a special case of the following general version of Bayes' rule

Fact (Bayes' Rule)**.** If $B_1, B_2, \ldots, B_m$ partition the sample space Ω, then

$$P\{B_k|A\} = \frac{P\{A|B_k\}P\{B_k\}}{\sum_{i=1}^{m} P\{A|B_i\}P\{B_i\}}$$

for any $k = 1, 2, \ldots, m$.

If we now let X and Y be discrete random variables and assume that $y_1, y_2, \dots, y_m$ are the possible values of Y, then with $A = \{X = x\}$ and $B_i = \{Y = y_i\}$ we conclude

$$P\{X = x\} = \sum_{i=1}^{m} P\{X = x, Y = y_i\} = \sum_{i=1}^{m} P\{X = x|Y = y_i\} P\{Y = y_i\}.$$

Note that this also works if Y takes on countably many values.

Example. Suppose that N_1 and N_2 are independent random variables with

$$P\{N_i = n\} = \frac{1}{2^{n+1}}, \quad n = 0, 1, 2, \dots.$$

Compute $P\{N_1 + N_2 = 7\}$.

Solution. Using the notation above, if we let $X = N_1 + N_2$, $Y = N_2$, $x = 7$, and $y_i = i$, then

$$\begin{aligned} P\{N_1 + N_2 = 7\} = \sum_{i=0}^{7} P\{N_1 + N_2 = 7, N_2 = i\} &= \sum_{i=0}^{7} P\{N_1 = 7 - i, N_2 = i\} \\ &= \sum_{i=0}^{7} P\{N_1 = 7 - i\} P\{N_2 = i\} \end{aligned}$$

where the last equality follows from the independence of N_1 and N_2. Thus,

$$\sum_{i=0}^{7} P\{N_1 = 7 - i\} P\{N_2 = i\} = \sum_{i=0}^{7} \frac{1}{2^{8-i}} \cdot \frac{1}{2^{i+1}} = \frac{8}{2^9} = \frac{1}{64}.$$

In the discrete case, we have seen that the law of total probability is

$$P\{X = x\} = \sum_{i=1}^{m} P\{X = x, Y = y_i\} = \sum_{i=1}^{m} P\{X = x|Y = y_i\} P\{Y = y_i\}.$$

In the continuous case, we know that $P\{X = x\} = 0$ for every value of x. Nonetheless, the discrete case *motivates* the following definition.

Definition. Let $(X, Y)'$ be continuous. If $f_Y(y) > 0$, then the *conditional density function of X given $Y = y$* is

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

Thus, the law of total probability for continuous random variables takes the form

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, \mathrm{d}y = \int_{-\infty}^{\infty} f_{X|Y=y}(x) f_Y(y) \, \mathrm{d}y.$$