Notion of Conditional Probabilities: Given two events X and Y that can occur in the random experiment, then $P[Y|X]$ is the probability of Y occurring given that the event X has occurred.

The conditional probability results in a refinement of the probability of Y based on new information from observations of X.

Conduct the experiment n times; Let $n_X$ be the number of times $X = x$ occurs and $n_{xy}$ be the number of times both $X = x$, $Y = y$ occur.

Then the ratio $P[Y|X] = \frac{n_{xy}}{n_x}$ represents the conditional probability.

Divide the RHS numerator and denominator by n and note that

$$\frac{n_{xy}}{n} = \frac{n_{xy}}{n_x} = \frac{P[X = x, Y = y]}{P[X]} = \frac{P[X, Y]}{P[X]}$$
Therefore the conditional probability can be represented as the ratio of the Joint Probability $P[X, Y]$ and the Marginal (or Total) Probability $P[X]$

$$P[Y|X] = \frac{P[X, Y]}{P[X]}$$

(1)

Note: $P[X] = \sum_Y P[X, Y] = \sum_Y P[X|Y] P[Y]$: Marginal Probability Distribution

- Where $P[X, Y] = P[Y|X] P[X] = P[X|Y] P[X]$ is referred to as the Product Rule
- And $P[X] = \sum_Y P[X, Y]$ is referred to as the Sum Rule

Notation: We use the notation $P[X], P[Y]$ to denote both probabilities and distributions. Refer to the context to interpret.

- When probability of $X = x$ is evaluated, $P[X] = P[X = x]$
- Otherwise $P[X]$ is a function: Probability mass function (PMF), $P[X = x_i]$ $i = 0, 1, ...$
Bayes Theorem

- Consider an outcome $Y$ of an experiment that depends on $n$ mutually exclusive events $X_1, X_2, \ldots X_n$.

- The probability $P(Y)$ referred as the total probability can be composed from its dependence on each of the $X_i$ as follows:

$$P(Y) = \sum_{i=1}^{n} P(Y, X_i) = \sum_{i=1}^{n} P(Y|X_i) P(X_i) \quad (2)$$

- where $P(Y|X_i)$ is the probability of $Y$ resulting from observation of $X_i$

- and $P(X_i)$ are known as the causal or *apriori* probabilities

- From these measurements, the *aposteriori* probability is given by Bayes Rule:

$$P(X_i|Y) = \frac{P(Y|X_i) P(X_i)}{P(Y)} \quad (3)$$
• [Summary: Bayes Theorem links prior information with the results of new observations and provides a mechanism to refine the prior probabilities]
Example: Consider the TCP data we’ve been working with which has attributes: (time-stamp, src-port, dest-port, pktSize)

Let $X$ represent the $\text{pktSize}$ as a RV: Range $x : (0, 1500)$ Bytes  
Let $Y$ represent the $\text{srcport}$ as a RV: Range $y : (0, xxxxx)$  
Let $Z$ represent the $\text{destport}$ as a RV: Range $z : (0, xxxxx)$

- Then (srcport, destport) values can be mapped to an application/protocol type Ex: [http-server: 80, smtp (email): xx, FTP : yy .... ]
- Probabilities $P(X)$, $P(Y)$, $P(Z)$ can be estimated from the data
- Of interest are conditional probabilities: $P(Y|X)$, $P(Z|X)$