 Hao Zhang nhz@eecs.berkeley.edu

Disclaimer: *These notes have not been subjected to the usual scrutiny for formal publications. They are to be used only for the class.*

Outline:

1. Terms in Probability
2. Bayes Rule/Chain Rule
3. misc(variance, binomial coeffs)

1 Terms in Probability

A simple example is

Prob of X,Y	X_1	X_2
Y_1	0.1	0.2
Y_2	0.4	0.3

$$\begin{aligned}
&& \textsc{Joint Distribution:} \\
\Pr[X = X_1, Y = Y_2] &=& 0.4 \\
\Pr[X = X_2, Y = Y_1] &=& ? \\
&& \textsc{Marginal Distribution:} \\
\Pr[X = X_1] &=& \Pr[X = X_1 \wedge (Y = Y_1 \vee Y = Y_2)] \\
&=& \Pr[(X = X_1 \wedge Y = Y_1) \vee (X = X_1 \wedge Y = Y_2)] \\
&=& \Pr[X = X_1, Y = Y_1] + \Pr[X = X_1, Y = Y_2] \\
&=& 0.5 \\
\Pr[X = X_2] &=& ? \\
&& \textsc{Conditional Probability:} \\
\Pr[X = X_1 | Y = Y_1] &=& \frac{\Pr[X = X_1, Y = Y_1]}{\Pr[Y = Y_1]} = \frac{0.1}{0.1 + 0.2} = \frac{1}{3} \\
\Pr[Y = Y_1 | X = X_1] &=& ?
\end{aligned}$$

In general, given joint distribution $\Pr[X = X_i, Y = Y_j]$, where $i = 1, 2, 3, ..., n$ and $j = 1, 2, 3, ..., m$. The *marginal distribution* is defined as $\Pr[X = X_i] = \sum_{j=1}^{m} \Pr[X = X_i, Y = Y_j]$. The *conditional probability* is defined as $\Pr[X = X_i | Y = Y_j] = \Pr[X = X_i, Y = Y_j] / \Pr[Y = Y_j]$. More generally, conditional probability can be defined on two events A and B:

$$\Pr[A|B] = \frac{\Pr[A \wedge B]}{\Pr[B]}$$

Note: The definitions above also hold for infinite discrete probability distribution, namely, when n and m are infinite.

2 Bayes Rule/Chain Rule

Question: Consider two events A, B, given the conditional probabilities $\Pr[A|B]$ and $\Pr[A|\bar{B}]$, and the marginal distribution $\Pr[B]$. Do we know about $\Pr[B|A]$?

A bit of derivation shows that the answer is yes:

$$\begin{aligned}\Pr[B|A] &= \frac{\Pr[B \wedge A]}{\Pr[A]} = \frac{\Pr[A|B]\Pr[B]}{\Pr[A \wedge B] + \Pr[A \wedge \bar{B}]} \\ &= \frac{\Pr[A|B]\Pr[B]}{\Pr[A|B]\Pr[B] + \Pr[A|\bar{B}]\Pr[\bar{B}]}\end{aligned}$$

Example (Power of prior): Let A, B be events: A = Tom has disease D and B = Tom is diagnosed(i.e. guessed) to have disease D. It is know that in Tom's community, only a portion of 10^{-4} people have the disease, i.e., $\Pr[A] = 10^{-4}$. However, the bad news is that the diagnose is pretty reliable, with probability 0.99. Let's model the diagnose as $\Pr[B|A] = 0.99$ and $\Pr[B|\bar{A}] = 0.05$ (the false-alarm rate). (Note that the last two numbers do not have to add up to 1)

The question is: if Tom is diagnosed to have the disease, how likely is he actually sick? that is, we want to know $\Pr[A|B]$.

Using Bayes rule,

$$\begin{aligned}\Pr[A|B] &= \frac{\Pr[B|A]\Pr[A]}{\Pr[B|A]\Pr[A] + \Pr[B|\bar{A}]\Pr[\bar{A}]} \\ &= \frac{0.99 \times 10^{-4}}{0.99 \times 10^{-4} + 0.05 \times (1 - 10^{-4})} \approx \frac{10^{-4}}{10^{-4} + 0.05} \approx \frac{10^{-4}}{0.05} = 0.002\end{aligned}$$

$\Pr[A]$ is called the "prior".
$\Pr[A|B]$ is called the "posterior".
So, we see that the small prior leads to the small posterior. Tom is unlikely to have the disease.

By definition,

$$\begin{aligned}\Pr[A, B] &= \Pr[A|B]\Pr[B] = \Pr[B|A]\Pr[A] \\ \Pr[A, B, C] &= \Pr[A]\Pr[B|A]\Pr[C|A, B]\end{aligned}$$

The second equality is proved by expanding the right hand side by the definition of conditional probability. Generally, for events $X_1, X_2, ..., X_n$, the following *chain rule* holds:

$$\Pr[X_1, X_2, ..., X_n] = \Pr[X_1]\Pr[X_2|X_1] ... \Pr[X_n|X_1, X_2, ..., X_{n-1}]$$

3 misc topics

The variance of a random variable X is $Var[X] = E[(X - E[X])^2]$.

Binomial sum: $\sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} = (a + b)^n$. Think about filling n empty space with k balls of type a and $n - k$ balls of type b.