Week 4 Module

Probability - continued

Two-Way Distributions

What we have seen thus far have been primarily univariate distribution. Now consider multivariate distributions such as:

- what is the probability that you are dealt a card that is a face card *and* a spade?, formally this is known as a joint distribution $p(\text{card} = \text{face}, \text{suit} = \text{spade})$.

Or consider the table below (from the textbook) that contains hair color and eye color:

	Hair Color				
Eye Color	Black	Brunette	Red	Blond	Marginal (eye color)
Brown	0.11	0.20	0.04	0.01	0.37
Blue	0.03	0.14	0.03	0.16	0.36
Hazel	0.03	0.09	0.02	0.02	0.16
Green	0.01	0.05	0.02	0.03	0.11
Marginal (hair color)	0.18	0.48	0.12	0.21	1.0

and calculate the following quantities:

- $p(\text{h} = \text{blond})$

- $p(\text{e} = \text{blue})$

- $p(\text{h} = \text{blond}, \text{e} = \text{blue})$

- $p(\text{h} = \text{blond}) = p(\text{h} = \text{bl.}, \text{e} = \text{blue}) + p(\text{h} = \text{bl.}, \text{e} = \text{brown}) + p(\text{h} = \text{bl.}, \text{e} = \text{hazel}) + p(\text{h} = \text{bl.}, \text{e} = \text{green})$, this is known as a *marginal* distribution that is obtained my marginalizing or integrating out the other variable.

- Formally, the *marginal* distribution can be defined as:
    - $p(x) = \sum_y p(x, y)$ in the discrete case
    - $p(x) = \int p(x, y) dy$

Conditional Probability

With multivariate distributions, researchers are often interested in the distribution of one variable given a specific value of another variable. For instance consider,

- The probability that it is snowing at Bridger Bowl vs.
- The probability that it is snowing at Bridger Bowl, given that it is cloudy in Bozeman.

This is known as a conditional probability and will be an essential element in Bayesian data analysis.

QUESTION: using the hair - eye color table, answer the following questions:

- What is the probability of a person having red hair:

- What is the probability of a person having blue eyes:

- What is the probability of a person having blue eyes and red hair:

- What is the probability of a person having blue eyes given that they have red hair:

- What is the probability of a person having red hair given that they have blue eyes:

Thus far the questions were answered with intuition, how can we mathematically formulate conditional probability for two events, A and B?

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

In other words, the conditional probability is the ratio of the joint probability and the marginal probablity of the event we are conditioning on.

- This can also be sketched out using a Venn diagram.

Ch.5 Bayes Rule

Compare the two probability statements:

- P[1 foot (or more) of powder at Bridger]
- P[1 foot (or more) of powder at Bridger | blue light on Baxter Hotel]

The first probability statement consider two possible outcomes: 1 foot (or more) of new snow and less than 1 foot of new snow. The second probability statement incorporates some additional information (data) into the probability statement, namely that 2 or more inches of snow has fallen.

Bayes rule is the mathematical foundation for re-allocating credibility (or probability) when conditioning on data.

Bayes Rule and Conditional Probability

- Recall: the conditional probability $P(A|B) = \frac{P(A \cap B)}{P(B)}$. From here we do some algebra to obtain Bayes rule.

$$P(A|B) \times P(B) = \frac{P(A \cap B)}{P(B)} \times P(B) \tag{1}$$

$$P(A|B)p(B) = P(B|A)P(A) \tag{2}$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{3}$$

$$P(A|B) = \frac{P(B|A)P(A)}{\sum_A' P(B|A')P(A')} \tag{4}$$

- Either of the last two equations are called **Bayes Rule**, named after Thomas Bayes.

Bayes Rule with two-way discrete table

The conditional probability $p(c|r)$ is the joint probability $p(r,c)$ in the cell divided by the marginal probability $p(r)$

Recall the following two-way table:

	Hair Color				
Eye Color	Black	Brunette	Red	Blond	Marginal (eye color)
Brown	0.11	0.20	0.04	0.01	0.37
Blue	0.03	0.14	0.03	0.16	0.36
Hazel	0.03	0.09	0.02	0.02	0.16
Green	0.01	0.05	0.02	0.03	0.11
Marginal (hair color)	0.18	0.48	0.12	0.21	1.0

Previously we calculated:

- What is the probability of a person having red hair given that they have blue eyes

    We now see that this is a simple illustration of Bayes rule.

A classic example of Bayes rule focuses on diagnosing a rare disease. There are a few important values we need to state:

- Let θ be a parameter that determines whether a person has the disease,

- Let T be the result of the diagnostic test,

- Let $Pr(\theta = Yes) = p_\theta$ be the probability a person from the general population has the disease,

- Let $Pr(Test = Yes|\theta = Yes) = p_{T+}$ be the probability of detecting a disease when it is present. This is called the hit rate in the textbook,

- Let $Pr(Test = Yes|\theta = No) = p_{T-}$ be the probability of a false alarm.

- **Question:** do we need to state,
    - $Pr(\theta = No)$
    - $Pr(Test = No|\theta = Yes)$
    - $Pr(Test = No|\theta = No)$

Assume we are testing citizens for Extra Sensory Perception (ESP). The ultimate goal will be to determine the probability that an individual has ESP if they test positive for ESP. Mathematically this is stated as $Pr(\theta = Yes|Test = Yes)$.

First using the generic probability from the previous page, compute $Pr(\theta = Yes|Test = Yes)$.

$$
\begin{aligned}
Pr(\theta = Yes|Test = Yes) &= \frac{Pr(Test = Yes|\theta = Yes)Pr(\theta = Yes)}{\sum_{\theta}' Pr(Test = Yes|\theta = \theta')Pr(\theta = \theta')} \\
&= \frac{Pr(Test = Yes|\theta = Yes)Pr(\theta = Yes)}{Pr(Test = Yes|\theta = Yes)Pr(\theta = Yes) + Pr(Test = Yes|\theta = No)Pr(\theta = No)} \\
&= \frac{p_{T+} \times p_{\theta}}{p_{T+} \times p_{\theta} + p_{T-} \times (1 - p_{\theta})}
\end{aligned}
$$

Now to make this concrete assume:

- The rate of ESP in the population is 1 in 100,000, so $Pr(\theta = Yes) = p_{\theta} = 0.00001$

- The hit rate of the test is 9999 in 10000 %, so $Pr(Test = Yes|\theta = Yes) = p_{T+} = .9999$

- The false detection rate is 1 in 10000 %, so $Pr(Test = Yes|\theta = No) = p_{T-} = 0.0001$

- **Question:** Before doing any math, what is your guess for the probability that a person receiving a positive test actually has ESP?

```
p.theta <- 1 / 100000
p.t.plus <- 9999 / 10000
p.t.minus <- 1 / 10000
p.theta.true <- p.t.plus * p.theta / (p.t.plus * p.theta + p.t.minus * (1 - p.theta))
```

It turns out that the probability that a person actually has ESP given they had a positive test is $Pr(\theta = Yes|Test = Yes) = 0.0909$.

This example allows us to understand the mechanisms behing Bayes rule.

- ESP is rare (if you believe in that kind of thing), so there is low prior probability of a person having this ability

- The test provides some additional information to support the person having ESP - that is there is a shift in the plausibility of this outcome, but the combination of the rareness of ESP and the false detection rate still make it more likely that the person does not have ESP.

Bayes rule with parameters and data

The previous example was essentially a probability exercise and we were not doing Bayesian statistical analysis per se, but rather just using Bayes rule. Bayesian statistical analysis refers to a fairly specific application of this theorem where:

- there is a statistical model of observed data, conditional on parameter values, $p(\mathcal{D}|\theta)$. We will use $\mathcal{D}$ to represent data and θ to represent the parameters. (Note the statistical model of observed data exists in a frequentist paradigm as well, as this function is sometimes called a *statistical likelihood* which is used for Maximum Likelihood Estimation (MLE).)

- there exists a **prior** belief on the parameter values, $p(\theta)$, and

- Bayes rule is used to convert the prior belief on the parameters *and* the statistical model into a **posterior** belief $p(\theta|\mathcal{D})$.

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})}$$

The denominator $p(\mathcal{D})$ is referred to as the marginal likelihood of the data and is computed as:

- $p(\mathcal{D}) = \sum_{\theta'} p(\mathcal{D}|\theta')p(\theta')$ in the case where the parameters are discrete and $p(\mathcal{D}) = \int p(\mathcal{D}|\theta')p(\theta')d\theta'$ in the case where the parameters are continuous. The θ' is used as we enumerate or integrate across all possible values of the parameter values.