Conditional Probability

STAT 211 - 509

2018-09-18

Conditional Probability

Joint Distributions

Consider two discrete random variables X and Y with probability functions f_X and f_Y, respectively. Let $f_{X,Y}$ be the **joint probability function** for X and Y.

- $$f_{X,Y}(x, y) \geq 0 \quad \forall (x, y)$$
- $$\sum_x \sum_y f_{X,Y}(x, y) = 1$$
- For any set of possible values A

$$P((X, Y) \in A) = \sum_{(x,y) \in A} f_{X,Y}(x, y)$$

This extends to 3 or more random variables considered jointly.

Marginal Distributions

The **marginal probability functions** for discrete X and Y are defined by

$$f_X(x) = \sum_y f_{X,Y}(x, y)$$

$$f_Y(y) = \sum_x f_{X,Y}(x, y)$$

By summing over the values of one variable, we recover the "marginal" probability function for the other.

Conditional Probability

For two random variables X and Y the **conditional probability** that $Y = y$ given that $X = x$ is

$$\begin{aligned} f_{Y|X}(y|x) &= P(Y = y|X = x) \\ &= \frac{P(X = x, Y = y)}{P(X = x)} \\ &= \frac{f_{X,Y}(x,y)}{f_X(x)} \end{aligned}$$

Think of this probability as the proportion of times that Y takes the value y among those times for which $X = x$.

Example: How Couples Meet and Stay Together

- Data from Interuniversity Consortium for Political and Social Research (ICPSR)
  - `http://` www.icpsr.umich.edu/icpsrweb/ICPSR/studies/30103
- Data File and R Code:
  - HCMST.csv
  - HCMST.r
- Things to Try:
  - Compare GLB status to political party affiliation.
  - Among these 3,009 individuals:
    * What is the joint distribution, what are the marginal distributions, and what is the conditional distribution of political party affiliation, given GLB status? http:// www.icpsr.umich.edu/icpsrweb/ICPSR/studies/30103

Independence Revisited

The two random variables X and Y are **independent** if

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

Under independence, we have

$$\begin{aligned} f_{Y|X}(y|x) &= \frac{f_{X,Y}(x,y)}{f_X(x)} \\ &= \frac{f_X(x) f_Y(y)}{f_X(x)} \\ &= f_Y(y) \end{aligned}$$

That is, the probabilities that govern Y do not depend on the value taken by X.

Example

In the HCMST sample population, are GLB status and political party affiliation independent?

Law of Total Probability

With S_X the sample space for the random variable X, let $A_1, A_2, \ldots, A_k$ be a **partition** of S_X:

$$A_1 \cup A_2 \cup \ldots \cup A_k = S_X$$

with $P(A_i) > 0,\ i = 1, 2, \ldots, k$.

The **Law of Total Probability** states that:

$$P(Y = y) = \sum_{i=1}^{k} P(Y = y | A_i) P(A_i)$$

Bayes' Theorem

Consider again the partition $\{A_i\}$. **Bayes' Theorem** states that:

$$P(A_i | Y = y) = \frac{P(Y = y | A_i) P(A_i)}{\sum_j P(Y = y | A_j) P(A_j)}$$

Notice that Bayes' Theorem allows us to switch from conditional probabilities for Y to conditional probabilities for A_i.

$P(A_i)$ is called the **prior probability** of A_i, and $P(A_i | Y = y)$ is called the **posterior probability** of A_i.

Example

In the HCMST sample population:

- Apply the law of total probability to compute the probability of democrat.
- Apply Bayes' Theorem to compute the probability of GLB, given democrat.

Example

- Setting: Anytown , USA. Population: 200,001.
- Robbery: Perpetrator described as being a male, 20-25 years old, 6'10" tall, red hair, with a limp. Only 5 people fitting that description in the city.
- Man who fits description spotted nearby and arrested.
- Prosecutor: The chances that we would find an innocent man nearby who happened to fit the description is too small to believe.

Example Continued

We have that

$$P(\text{fits description}|\text{not guilty}) = \frac{4}{200000} = 0.00002$$

a very small probability. But:

$$\begin{aligned} P(\text{not guilty}|\text{fits description}) &= \frac{P(\text{fits description}|\text{not guilty})P(\text{not guilty})}{P(\text{fits description})} \\ &= \frac{\frac{4}{200000}\frac{200000}{200001}}{\frac{5}{200001}} \\ &= \frac{4}{5} = 0.8 \end{aligned}$$

This is an illustration of the *prosecutor's fallacy* .

See the work of Distinguished Professor Cliff Spiegelman (Dept. of Statistics, TAMU) for examples of how statistical thinking is impacting the legal system in the USA.

Conditional Expectation

The conditional expectation of the discrete random variable X given that $Y = y$ is

$$E[X|Y = y] = \sum_x x f_{X|Y}(x|y)$$

Let g be a function $\mathbb{R}^2 \to \mathbb{R}$

$$E[g(X,Y)|Y = y] = \sum_x g(x,y) f_{X|Y}(x|y)$$

Law of total expectation:

$$E[E(Y|X)] = E[Y]$$
$$E[E(X|Y)] = E[X]$$

Example

- Data on 261 students in an undergraduate introductory statistics class.
- Data File and R Code:
  - stat_class.csv
  - stat_class.r
- Among these 261 students:

- What is the expected score on the final exam, conditional on 10 or more absences?
- What is the expected score on the final exam, conditional on fewer than 10 absences?